

Genfore version 4.5
Copyright (c) 1994 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2002, 15:33:02 : Search time 19.99 seconds
(without alignments)

2619.744 Million cell updates/sec

Files: US-09-603-665-5

Perfect score: 2144

Sequences: 1 MISAQIQLRELALGSHASL.....CQRTIQQLLEIVIGEPHQSVP 2144

Scoring table: OLF30

Gapop 60.0 : Gapext 60.0

Searched: 233628 seqs, 2442594 residues

Word size : 6

Total number of hits satisfying chosen parameters: 2393

Minimum DB seq length: 0

Maximum DB seq length: 2040000000

Post-processing: listing first 45 summaries

Database : Issued_Patents_AA *

- 1: 2002/6/prod/1/2002/6/AA_0006_P1 *
- 2: 2002/6/prod/1/2002/6/AA_0006_P2 *
- 3: 2002/6/prod/1/2002/6/AA_0006_P3 *
- 4: 2002/6/prod/1/2002/6/AA_0006_P4 *
- 5: 2002/6/prod/1/2002/6/AA_0006_P5 *
- 6: 2002/6/prod/1/2002/6/AA_0006_P6 *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.4	144	1	US-08-284-393B-11
2	8	0.4	144	1	US-08-259-696B-9
3	8	0.4	144	2	US-08-759-628-6
4	8	0.4	144	4	US-09-522-217-114
5	8	0.4	144	5	US-09-522-217-114
6	8	0.4	829	4	US-09-413-814-105
7	8	0.4	12	3	US-08-740-069-6
8	7	0.3	12	3	US-09-282-501-6
9	7	0.3	12	4	US-08-610-915A-79
10	7	0.3	71	4	US-08-905-223-311
11	7	0.3	71	4	US-08-905-223-311
12	7	0.3	38	1	US-08-211-202-130
13	7	0.3	117	3	US-08-545-809A-139
14	7	0.3	119	2	US-08-475-000-16
15	7	0.3	119	2	US-08-483-199-16
16	7	0.3	119	2	US-08-484-508-16
17	7	0.3	143	1	US-08-259-696B-10
18	7	0.3	143	2	US-08-502-514-10
19	7	0.3	144	1	US-08-133-979A-11
20	7	0.3	144	2	US-08-436-890-11
21	7	0.3	144	2	US-08-451-213-11
22	7	0.3	178	2	US-08-928-692-57
23	7	0.3	179	1	US-08-665-966-8
24	7	0.3	179	3	US-09-041-780-8
25	7	0.3	214	1	US-08-217-327-4
26	7	0.3	238	3	US-09-079-981-1
27	7	0.3	432	1	US-08-961-083-12

28	7	0.3	315	1	US-07-729-099-1	Sequence 1, Appl
29	7	0.3	315	1	US-08-257-392-1	Sequence 1, Appl
30	7	0.3	315	3	US-08-770-035-1	Sequence 1, Appl
31	7	0.3	331	3	US-08-556-419-25	Sequence 25, Appl
32	7	0.3	343	3	US-08-978-741-17	Sequence 17, Appl
33	7	0.3	343	4	US-09-343-729A-13	Sequence 13, Appl
34	7	0.3	344	1	US-08-843-994-4	Sequence 3, Appl
35	7	0.3	344	3	US-09-054-520A-3	Sequence 3, Appl
36	7	0.3	344	3	US-09-334-275-3	Sequence 3, Appl
37	7	0.3	365	3	US-08-978-741-2	Sequence 2, Appl
38	7	0.3	365	4	US-09-334-729A-3	Sequence 3, Appl
39	7	0.3	392	1	US-08-723-948-1	Sequence 1, Appl
40	7	0.3	392	2	US-09-080-538-1	Sequence 1, Appl
41	7	0.3	397	3	US-08-978-741-6	Sequence 6, Appl
42	7	0.3	397	4	US-03-333-729A-7	Sequence 7, Appl
43	7	0.3	428	1	US-09-347-833-6	Sequence 6, Appl
44	7	0.3	446	1	US-08-665-966-10	Sequence 10, Appl
45	7	0.3	446	3	US-09-041-780-10	Sequence 10, Appl

ALIGNMENTS

```

RESULT 1
US-08-284-393B-11
: Sequence 11, Application US/08284393B
: Patent No. 5696234
: GENERAL INFORMATION:
: APPLICANT: Zurawski, Sandra M.
: APPLICANT: Zurawski, Gerard
: TITLE OF INVENTION: METHODS OF MAMMARY CANCER TREATMENT
: NUMBER OF SEQUENCES: 16
: REFERENCE SEQUENCE ACCESS:
: ADDRESSEE: DNAX Research Institute
: STREET: 901 California Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304-1104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.40
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/284,393B
: FILING DATE: 01-AUG-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Ching, Edwin P.
: REGISTRATION NUMBER: 34,090
: REFERENCE/WORKSHEET NUMBER: DX0389
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-852-9196
: TELEFAX: 415-496-1200
: INFORMATION FOR SPO ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 144 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
US-08-284-393B-11

```

Query Match 0.43 Score 8; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8, Conservative 0; Mismatches 0; Gaps 0;

QY 1305 LLLLGTV 1312

IIIIIIII

DB 6 LLLLGTV 13

RESULT 2
US-08-259-696B 9
Sequence 9, Application US/08259696B
Patent No. 5702919
GENERAL INFORMATION:
APPLICANT: Nash, Richard A.
TITLE OF INVENTION: DNA ENCODING CANINE GRANULOCYTE
TITLE OF INVENTION: MACROPHAGE COLONY STIMULATING FACTOR (as amended)
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/259,696B
FILING DATE: 14-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: FUSP17718
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-224-0718
TELEFAX: 206-224-0779
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: Human GMCSE, Figure 5
US-08-259-696B-9

Query Match 0.4%; Score 8; DB 1; Length 144,
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 LLLLGTV 1312
DB 6 LLLLGTV 13

RESULT 3
US-08-902-513-9
Sequence 9, Application US/08902513
Patent No. 5925343
GENERAL INFORMATION:
APPLICANT: Nash, Richard A.
TITLE OF INVENTION: CANINE GRANULOCYTE MACROPHAGE
TITLE OF INVENTION: COLONY STIMULATING FACTOR
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/902,513
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/259,696
FILING DATE: 14-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: FUSP11036
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-224-0718
TELEFAX: 206-224-0779
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: Human GMCSE, Figure 5
US-08-902-513-9

Query Match 0.4%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 LLLLGTV 1312
DB 6 LLLLGTV 13

RESULT 4
US-08-759-628-6
Sequence 6, Application US/08759628
Patent No. 6225446
GENERAL INFORMATION:
APPLICANT: Altmann, Scott W.
APPLICANT: Rock, Fernando L.
APPLICANT: Baran, J. Fernando
TITLE OF INVENTION: MUTATIONAL VARIANTS OF MAMMALIAN PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,628
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,574
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0552Q
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-852-9196
 TELEFAX: 415-496-1200
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 144 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FEATURES:
 NAME/KEY: peptide
 LOCATION: 14..144
 OTHER INFORMATION: /note= "peptide of Figure 1"

US 08 759 628 6

Query Match 0.4% Score 8; DB 4; Length 144;
 Best Local Similarity 100.0%; Prod. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 LLLLGIVA 1312
 |||||
 Db 6 LLLLGIVA 13

RESULT 5
 US-09-522 217 114
 : Sequence 114, Application US/09/522217
 : Patent No. 6407024
 : GENERAL INFORMATION:
 : APPLICANT: No. 6407024ak, Julia E.
 : APPLICANT: Presnell, Scott R.
 : APPLICANT: Sprecher, Cindy A.
 : APPLICANT: Foster, Donald C.
 : APPLICANT: Holly, Richard D.
 : APPLICANT: Gross, Jane A.
 : APPLICANT: Johnston, Janet V.
 : APPLICANT: Nelson, Andrew J.
 : APPLICANT: Dillon, Stacy K.
 : APPLICANT: Hammond, Angela K.
 : TITLE OF INVENTION: Novel Cytokine ZALPHALL LIGAND
 : FILE REFERENCE: 99-16
 : CURRENT APPLICATION NUMBER: US/09/522,217
 : EARLIER FILING DATE: 2000-03-09
 : EARLIER APPLICATION NUMBER: US 60/123,547
 : EARLIER FILING DATE: 1999-03-09
 : EARLIER APPLICATION NUMBER: US 60/123,904
 : EARLIER FILING DATE: 1999-03-11
 : EARLIER APPLICATION NUMBER: US 60/142,013
 : EARLIER FILING DATE: 1999-07-01
 : NUMBER OF SEQ ID NOS: 115
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 114
 : LENGTH: 144
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-09-522-217-114

Query Match 0.4% Score 8; DB 4; Length 144;
 Best Local Similarity 100.0%; Prod. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1405 LLLLGIVA 1312
 |||||
 Db 6 LLLLGIVA 13

RESULT 6
 PCT-US95-08950 11
 : Sequence 11, Application PCT/US9508950
 : GENERAL INFORMATION:
 : APPLICANT: Zurawski, Sandra M.

APPLICANT: Zurawski, Gerard
 TITLE OF INVENTION: MUTEINS OF MAMMALIAN CYTOKINES
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DNAX Research Institute
 STREET: 901 California Avenue
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PC /US95/08950
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/284,393
 FILING DATE: 01-AUG-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Ching, Edwin P.
 REGISTRATION NUMBER: 34,090
 REFERENCE/DOCKET NUMBER: DX0389
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-852-9196
 TELEFAX: 415-496-1200
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 144 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-08950-11

Query Match 0.4% Score 8; DB 5; Length 144;
 Best Local Similarity 100.0%; Prod. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 LLLLGIVA 1312
 |||||
 Db 6 LLLLGIVA 13

RESULT 7
 US-09-413-814-105
 : Sequence 105, Application US/09413814
 : Patent No. 6225064
 : GENERAL INFORMATION:
 : APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
 : APPLICANT: Bristol-Myers Squibb, Co.
 : APPLICANT: Beyer, Stefan
 : APPLICANT: Bloecker, Helmut
 : APPLICANT: Brandt, Petra
 : APPLICANT: Cino, Paul M
 : APPLICANT: Dougherty, Brian A
 : APPLICANT: Goldberg, Steven L
 : APPLICANT: Hoffe, Gerhard
 : APPLICANT: Mueller, Joachim
 : APPLICANT: Reichenbach, Hans
 : TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
 : FILE REFERENCE: PCT/US 99/23535
 : CURRENT APPLICATION NUMBER: US/09/413,814
 : CURRENT FILING DATE: 1999-10-07
 : EARLIER APPLICATION NUMBER: DE 198 46 493.2
 : EARLIER FILING DATE: 1998-10-09
 : NUMBER OF SEQ ID NOS: 107
 : SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 105
LENGTH: 829
TYPE: PKT
ORGANISM: Sorangium cellulosum
US-09-413-814-105

Query Match 0.4%, Score 8; DB 4; Length 829;
Best Local Similarity 100.0%; Pred. No. 1202;
Matches 8; Conservative 0; Mismatches 0; Indels 0;

Qy 1757 LLSALAAL 1764
Db 309 LLSALAAL 316

RESULT 8
US-08-740-069-6
Sequence 6, Application US/08748069
Patent No. 6057290
GENERAL INFORMATION:
APPLICANT: Chihio FUKIAGE et al.
TITLE OF INVENTION: ANGIOGENESIS INHIBITOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/740,069
FILING DATE: October 24, 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:

INFORMATION FOR SEQ ID NO. 6.
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-740-069-6

Query Match 0.3%, Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 1476 EKEETIP 1482
Db 2 EKEETIP 8

RESULT 9
US-09-282-501-6
Sequence 6, Application US/09282501

Patent No. 6214800
GENERAL INFORMATION:
APPLICANT: Chihio FUKIAGE et al.
TITLE OF INVENTION: ANGIOGENESIS INHIBITOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/282,501
FILING DATE: April 9, 1999

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/740,069

FILING DATE: October 24, 1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/243,822

FILING DATE: February 3, 1999

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX:

TELEX:

INFORMATION FOR SEQ ID NO. 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 12 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide
US-09-282-501-6

Query Match 0.3%, Score 7; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 1476 EKEETIP 1482
Db 2 EKEETIP 8

RESULT 10
US-08-630-915A-79
Sequence 79, Application US/08630915A
Patent No. 6309820

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

APPLICANT: HOFFMAN, No. 6309820h

APPLICANT: KAY, Brian K

APPLICANT: FOWLER, Dana M.

APPLICANT: McCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

TITLE OF INVENTION: USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York


```

1 COUNTRY: USA
2 ZIP: 10036-2711
3 COMPUTER READABLE FORM:
4 MEDIUM TYPE: Floppy disk
5 COMPUTER: IBM PC compatible
6 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: Patent In Release #1.0, Version #1.30
8 CURRENT APPLICATION DATA:
9 APPLICATION NUMBER: US/08/640,915A
10 FILING DATE: 04-APR-1996
11 CLASSIFICATION: 536
12 ATTORNEY/AGENT INFORMATION:
13 NAME: MISTOCK, S. Leslie
14 REGISTRATION NUMBER: 18,872
15 REFERENCE/DOCKET NUMBER: 1101 174
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: (212) 790-9090
18 TELEFAX: (212) 869-8864/9741
19 TELEX: 66141 PENNIE
20 INFORMATION FOR SEQ ID NO: 79:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 41 amino acids
23 TYPE: amino acid
24 STRANDEDNESS:
25 TOPOLOGY: unknown
26 MOLECULE TYPE: Peptide
27 US-08-630-915A-79

```

```

Query Match 0.3% Score 7; DB 4; Length 41;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 ENDEPC 1988
Db 29 ENDEPC 35

```

```

RESULT 11
US-08-905-224-411
: Sequence 411, Application US/08/905,224
: Patent No. 6222929
: GENERAL INFORMATION:
: APPLICANT: Edwards, Jean-Baptiste D.
: APPLICANT: Lueltel, Aymeric
: APPLICANT: Lacroix, Bruno
: TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
: NUMBER OF SEQUENCES: 503
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 501 West Broadway
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92101-3505
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/905,224
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelson, Ned A.
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 255-8550
: TELEFAX: (619) 255-0176
: INFORMATION FOR SEQ ID NO: 311:
: SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 71 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: LINEAR
: MOLECULE TYPE: PROTEIN
: ORIGINAL SOURCE:
: ORGANISM: Homo Sapiens
: TISSUE TYPE: Brain
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 69-81
: IDENTIFICATION METHOD: Von Heijne matrix
: OTHER INFORMATION: score 6.7
: OTHER INFORMATION: seq FPKLLILGAMCSQ/AR
: US-08-905-223-311

```

```

Query Match 0.3% Score 7; DB 4; Length 71;
Best Local Similarity 100.0%; Pred. No. 102;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 151 LKINNSK 157
Db 39 LKINNSK 45

```

```

RESULT 12
US-08-211-202-130
: Sequence 130, Application US/08/211,202
: Patent No. 5565332
: GENERAL INFORMATION:
: APPLICANT: HOOGENDOORN, Hendricus Renatus Jacobus Mattheus
: APPLICANT: HAFFER, Michael
: APPLICANT: JESPER, Laurent Stephane Anne Therese
: APPLICANT: WINTER, Gregory Paul
: TITLE OF INVENTION: Production of chimeric antibodies - a
: TITLE OF INVENTION: combinatorial approach
: NUMBER OF SEQUENCES: 144
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David W. Clough, Marshall O'Toole Gerstein Murray &
: ADDRESSEE: Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25 (Hpo)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/211,202
: FILING DATE: 23-SEP-1992
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9120252.3
: FILING DATE: 23-SEP-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9120377.8
: FILING DATE: 25-SEP-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9206318.9
: FILING DATE: 24-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9206372.6
: FILING DATE: 24-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/GB92/00883
: FILING DATE: 15-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: David W. Clough
: REGISTRATION NUMBER: 36,107
: REFERENCE/DOCKET NUMBER: 28111/31960

```

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 130:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 98 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-211-202-130

Query Match 0.3%, Score 7; DB 1; Length 98;
 Best Local Similarity 100.0%, Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2036 KFORVMT 2042
 |||||
 DB 63 KFORVMT 69

RESULT 13
 US-08-545-809A-139
 Sequence 139, Application US/08545809A
 Patent No. 6096878
 GENERAL INFORMATION:
 APPLICANT: Honjo, Tasuku
 APPLICANT: Matsuda, Fumiko
 TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
 NUMBER OF SEQUENCES: 145
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson, P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: US
 ZIP: 02110 2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FASTSP0 for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/545.809A
 FILING DATE: 27-MAR-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP93/00603
 FILING DATE: 10-MAY-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Freeman, John W.
 REGISTRATION NUMBER: 29,066
 REFERENCE/DOCKET NUMBER: 06501/0004001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8906
 TELEX: 200154

INFORMATION FOR SEQ ID NO: 139:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 117 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-545-809A-139

Query Match 0.3%, Score 7; DB 3; Length 117;
 Best Local Similarity 100.0%, Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2036 KFORVMT 2042
 |||||

DB 82 KFORVMT 88

RESULT 14
 US-08-475-000-16
 Sequence 16, Application US/08475000
 Patent No. 5811267
 GENERAL INFORMATION:
 APPLICANT: RING, DAVID B.
 TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CHIRON CORPORATION
 STREET: INTELLECTUAL PROPERTY B440 - P.O. BOX 8097
 CITY: Emeryville
 STATE: CA
 COUNTRY: USA
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/475,000
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: SAVERIDE, PAUL
 REGISTRATION NUMBER: 36,914
 REFERENCE/DOCKET NUMBER: 0850.007
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 601-2585
 TELEFAX: (510) 655-3542
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 119 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-475-000-16

Query Match 0.3%, Score 7; DB 2; Length 119;
 Best Local Similarity 100.0%, Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1533 KVVESGG 1539
 |||||
 DB 3 KVVESGG 9

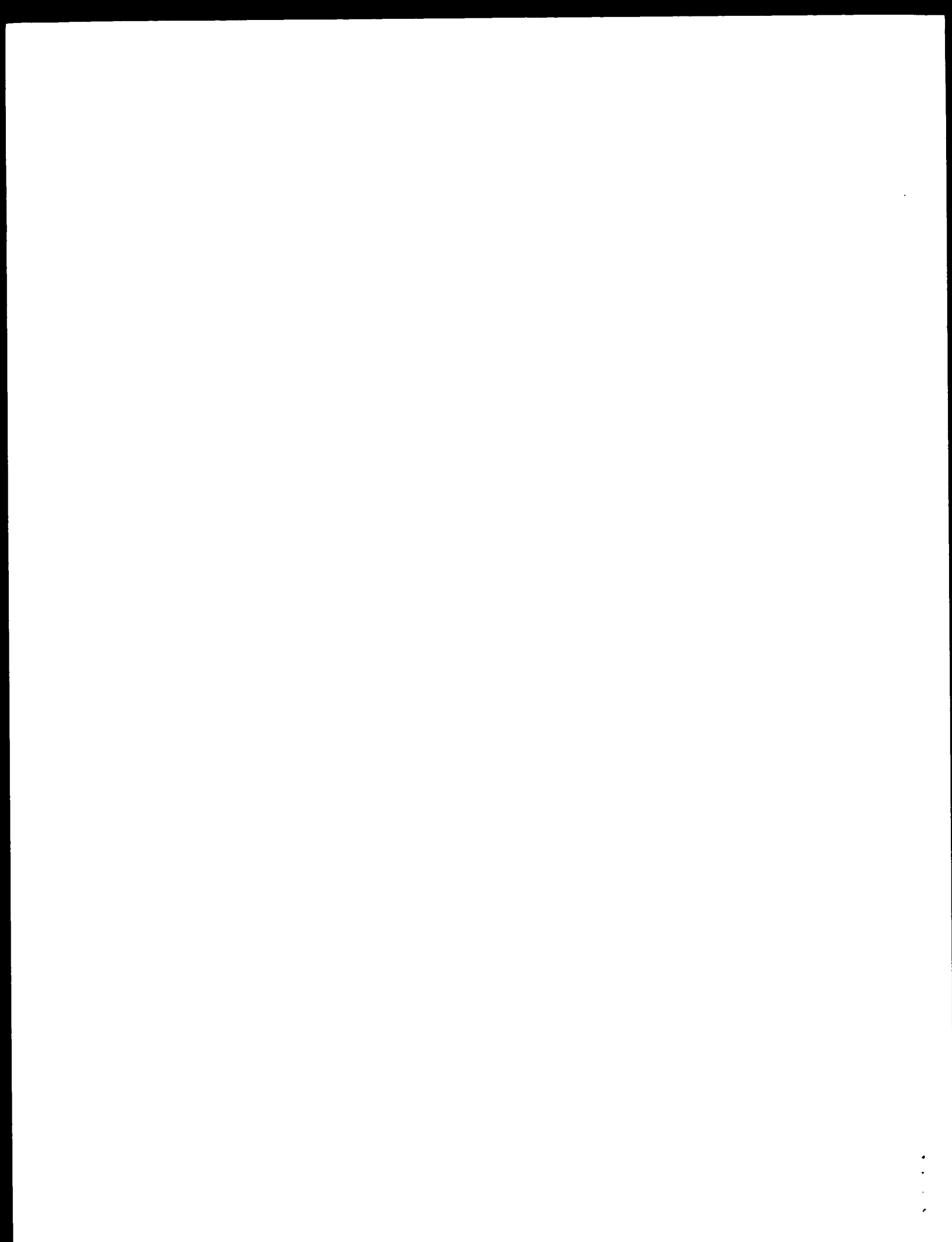
RESULT 15
 US-08-483-199-16
 Sequence 16, Application US/08483199
 Patent No. 5849877
 GENERAL INFORMATION:
 APPLICANT: RING, DAVID B.
 TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CHIRON CORPORATION
 STREET: INTELLECTUAL PROPERTY - B440, P.O. BOX 8097
 CITY: Emeryville
 STATE: CA
 COUNTRY: USA
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

1 SOFTWARE: Patent In Release #1.0, Version #1.25
 2 CURRENT APPLICATION DATA:
 3 APPLICATION NUMBER: US/09/483,199
 4 FILING DATE: 07-JUN-1995
 5 CLASSIFICATION: 424
 6 ATTORNEY/AGENT INFORMATION:
 7 NAME: SAVERIDE, PAUL
 8 REGISTRATION NUMBER: 36,914
 9 REFERENCE/DOCKET NUMBER: 0850,009
 10 TELECOMMUNICATION INFORMATION:
 11 TELEPHONE: (510) 601-2585
 12 TELEFAX: (510) 655-8542
 13 INFORMATION FOR SEQ ID NO: 16:
 14 SEQUENCE CHARACTERISTICS:
 15 LENGTH: 119 amino acids
 16 TYPE: amino acid
 17 TOPOLOGY: linear
 18 MOLECULE TYPE: protein
 19 US-08 483,199 16

Query Match 0.38; Score 7; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1543 KVVESRG 1549
 1111111
 16 4 KVVESRG 9

Search completed: July 2, 2002, 15:36:40
 Job time: 218 sec




```

FT VAR/ANI 2017 2017 /FTid=VAR_010941.
FT Rest local Similarity 99.94; /FTid=VAR_010942.
SQ SEQUENCE 2144 AA: 242455 MW: 666166740R7 0P64:

Query Match 82.3%; Score 1764; DB 1; Length 2144;
Rest local Similarity 99.94; Pred. No 0;
Matches 1964; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTSIAQQLKALPQSLASLSRUEVASLLPDPKEAATIDRTAFAGCTGLEELLGIDP 60
DB 1 MTSIAQQLKALPQSLASLSRUEVASLLPDPKEAATIDRTAFAGCTGLEELLGIDP 60

QY 61 SFEQFEAPLFSQAKTLEKSVQTRAVNKOLDENTSLFIHLSPVFLKPAQCKLEWLHHR 120
DB 61 SFEQFEAPLFSQAKTLEKSVQTRAVNKOLDENTSLFIHLSPVFLKPAQCKLEWLHHR 120

QY 121 FHIHLYNODSLIACVLPYHETRIFRVRIQIHLKINNSKRWFWLIPVKGSGVPLAKGTIIT 180
DB 121 FHIHLYNODSLIACVLPYHETRIFRVRIQIHLKINNSKRWFWLIPVKGSGVPLAKGTIIT 180

QY 181 HCYKDLGPMDFICSLVTKSVKFAEYFPGSSAQLRVLLAFYASTIVSALVAADVDSDNIIA 240
DB 181 HCYKDLGPMDFICSLVTKSVKFAEYFPGSSAQLRVLLAFYASTIVSALVAADVDSDNIIA 240

QY 241 KLFPIYQKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSILKDG 300
DB 241 KLFPIYQKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSILKDG 300

QY 301 LSCILVILQROKPSIGKPPHHCNVPDLITLHGISETVDSVPLLRMYLPHIIVVSIH 360
DB 301 LSCILVILQROKPSIGKPPHHCNVPDLITLHGISETVDSVPLLRMYLPHIIVVSIH 360

QY 361 HVTGETEKMDSQIVYKRHLEALLTKISLKNLDHLASLLFEEYISYSSQEMDSNKVSL 420
DB 361 HVTGETEKMDSQIVYKRHLEALLTKISLKNLDHLASLLFEEYISYSSQEMDSNKVSL 420

QY 421 LNEQFLPIRLLESKYPRTLDVVLEHKEADLKKQELFQVFSLSLSTSGGKYQFLADSD 480
DB 421 LNEQFLPIRLLESKYPRTLDVVLEHKEADLKKQELFQVFSLSLSTSGGKYQFLADSD 480

QY 481 TSIMLSLNLHPLAPVITLANNHKKIMKTSKGVDRSPFIKEAVIARLGDNDIDVLSAISA 540
DB 481 TSIMLSLNLHPLAPVITLANNHKKIMKTSKGVDRSPFIKEAVIARLGDNDIDVLSAISA 540

QY 541 PEIFKEHFSSEVTISNLNLFPQAEKSKNGEYEVILKIAADLILKEELISENDQLSNQV 600
DB 541 PEIFKEHFSSEVTISNLNLFPQAEKSKNGEYEVILKIAADLILKEELISENDQLSNQV 600

QY 601 VCLLPFVWINDDTESAEKIAIYLSKSGICSLHPLLRGWEALENVIKSTKPGKLIGVA 660
DB 601 VCLLPFVWINDDTESAEKIAIYLSKSGICSLHPLLRGWEALENVIKSTKPGKLIGVA 660

QY 661 NOKMTELLADNINLGDPSMIMKMDLISVGEFFSFNKKQKVTFRVILSVISCCSSILKE 720
DB 661 NOKMTELLADNINLGDPSMIMKMDLISVGEFFSFNKKQKVTFRVILSVISCCSSILKE 720

QY 721 THFPFAIRVFSLLQKIKKLESVITAVEIPSEWHLELMDRGPVELWAHYVEELNSTQR 780
DB 721 THFPFAIRVFSLLQKIKKLESVITAVEIPSEWHLELMDRGPVELWAHYVEELNSTQR 780

QY 781 VAVEDSVFLVFSLLKFIYALKAPKSPKGDITWNPPEQLKEDSRDYLHLLIGLFEMMLNCA 840
DB 781 VAVEDSVFLVFSLLKFIYALKAPKSPKGDITWNPPEQLKEDSRDYLHLLIGLFEMMLNCA 840

QY 841 DAVHFRVIMKFLIKVHLEIDVFOIPKPCSVIWTYGGSSLNPIANCSTKVTIQTQALVYGCAM 900
DB 841 DAVHFRVIMKFLIKVHLEIDVFOIPKPCSVIWTYGGSSLNPIANCSTKVTIQTQALVYGCAM 900

QY 901 ISSQKTPKKEHLASISSPVVTSLLNLINLSPVKEVFAALIQTLALSSSVASPEYLIITGHL 960
DB 901 ISSQKTPKKEHLASISSPVVTSLLNLINLSPVKEVFAALIQTLALSSSVASPEYLIITGHL 960

```

RESULT 2

BP28_MACHA

110 HP28_MACHA STANJARD; PRT; 958 AA.

```

QY 961 SKAPEITTSAAAYVVTQIATLPEELQPKKLSHQKLSHETANEISVSVSSYIAKIDMK 1020
DB 961 SKAPEITTSAAAYVVTQIATLPEELQPKKLSHQKLSHETANEISVSVSSYIAKIDMK 1020

QY 1021 VLOGVNGHMVLSQIDPMAPOQLIERTQKEPTAVIKIPAMVILHETLOKYNPPSVSLINRDPK 1080
DB 1021 VLOGVNGHMVLSQIDPMAPOQLIERTQKEPTAVIKIPAMVILHETLOKYNPPSVSLINRDPK 1080

QY 1081 SLIDFIKAVHHTIKELYAGMPITQITALEKTIKPPFAALISDEKVQOKLRLMLFDLLVNCN 1140
DB 1081 SLIDFIKAVHHTIKELYAGMPITQITALEKTIKPPFAALISDEKVQOKLRLMLFDLLVNCN 1140

QY 1141 SHCAQTVSSVFKGTSVNAEQVRILEPPEKAKPILTVQCKKKQKMOOKKSQULSSVQKVG 1200
DB 1141 SHCAQTVSSVFKGTSVNAEQVRILEPPEKAKPILTVQCKKKQKMOOKKSQULSSVQKVG 1200

QY 1201 GSYWQPVTLIIFELQHKKKIISPOQIIVPTIFNLISRCLEPIIDPOGNNMYTKQIISQIL 1260
DB 1201 GSYWQPVTLIIFELQHKKKIISPOQIIVPTIFNLISRCLEPIIDPOGNNMYTKQIISQIL 1260

QY 1261 NTCOKISPDGCKIPKDIIDREKFNVELIVQCTIPISMPQTHHIALILGTVAGTIPDKVI 1320
DB 1261 NTCOKISPDGCKIPKDIIDREKFNVELIVQCTIPISMPQTHHIALILGTVAGTIPDKVI 1320

QY 1321 HNIMSIFTFMGANVMKLDPTYSPQVINKIVMVIIPALLOSUSGUSIEVSRNVEEIVVKII 1380
DB 1321 HNIMSIFTFMGANVMKLDPTYSPQVINKIVMVIIPALLOSUSGUSIEVSRNVEEIVVKII 1380

QY 1381 SVFVDALPHVPEHREPIILVQIVDTLGAEPILWILILLLFPQYVTKTVIAAAYGKDAII 1440
DB 1381 SVFVDALPHVPEHREPIILVQIVDTLGAEPILWILILLLFPQYVTKTVIAAAYGKDAII 1440

QY 1441 EADTEFWFSVCCFEFSVQHQIQSLMNIILYLLKLPKEKEETIPKAVSNKSESQEEMLQVF 1500
DB 1441 EADTEFWFSVCCFEFSVQHQIQSLMNIILYLLKLPKEKEETIPKAVSNKSESQEEMLQVF 1500

QY 1501 NVETHTSKOLRHFKFLSVSPMSQLSSNNFLKKVVSOGGPREIILKGLERLITVLGYISA 1560
DB 1501 NVETHTSKOLRHFKFLSVSPMSQLSSNNFLKKVVSOGGPREIILKGLERLITVLGYISA 1560

QY 1561 VAQSMERNADKLTIVKFWALLSKAYDLDLKVNAIILPTETFTIPVIRGIVGNPIPSVPRKAL 1620
DB 1561 VAQSMERNADKLTIVKFWALLSKAYDLDLKVNAIILPTETFTIPVIRGIVGNPIPSVPRKAL 1620

QY 1621 DLINNKIQONTISWKKTIIVTRFLKLYPDLLAIVQPKKKEEPEEJAINPJTALYTIKILNKN 1680
DB 1621 DLINNKIQONTISWKKTIIVTRFLKLYPDLLAIVQPKKKEEPEEJAINPJTALYTIKILNKN 1680

QY 1681 FGAENPDPPVPLNTAVKLIAPERKEKNVLSGALLCITAEVSTLEALAIPOLPSPISLSL 1740
DB 1681 FGAENPDPPVPLNTAVKLIAPERKEKNVLSGALLCITAEVSTLEALAIPOLPSPISLSL 1740

QY 1741 LTTMKNSTSLVSSEVYLLSALAALOKVETLPHFISPYLEGILSQVILHLEKITSUMGCSAS 1800
DB 1741 LTTMKNSTSLVSSEVYLLSALAALOKVETLPHFISPYLEGILSQVILHLEKITSUMGCSAS 1800

QY 1801 QANIRLTSLKKTTLATLAPPVLLPAIKKTYKQIEKNKNKMHMGPFWSTIEGHTGKMKKEEL 1860
DB 1801 QANIRLTSLKKTTLATLAPPVLLPAIKKTYKQIEKNKNKMHMGPFWSTIEGHTGKMKKEEL 1860

QY 1861 TSHQSULTAFFLEALUFKAQUSENDEEVGKTENCIIDCLVAMVVKLSEVTFEPLFKLF 1920
DB 1861 TSHQSULTAFFLEALUFKAQUSENDEEVGKTENCIIDCLVAMVVKLSEVTFEPLFKLF 1920

QY 1921 DWAKTEDADKDELFPYNLADCTAEKIKGLFTIIPAGHIVKPPADTL 1966
DB 1921 DWAKTEDADKDELFPYNLADCTAEKIKGLFTIIPAGHIVKPPADTL 1966

```

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF176816; AAF00990.1; .

KW Hypothetical protein; Coiled coil.

FT DOMAIN 47 1426 COILED COIL (POTENTIAL).

SO SEQUENCE 1453 AA; 16994 MW; 03CBA02A64CF4139 CRC64;

Query Match 0.5%; Score 10; DB 1; Length 1453;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 LQKIKKLES 742

DB 956 LQKIKKLES 955

|||||

RESULT 4

BP28_DROME STANDARD: PRI: 2096 AA.

AC Q9VM75;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein (CG10805).

GN CG10805.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Euryarchaea; Neoptera; Eulipeter; Diptera; Brachyera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

FX MEDLINE=20196006; PubMed=7071132;

FA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

FA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

FA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

FA Sutton G.G., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,

FA Brannon R.C., Rogers Y.-H.C., Blazej P.G., Champe M., Pfeiffer H.D.,

FA Wan K.H., Doyle C., Baxter E.G., Helt G., Helt G., Nikles G.L.G.,

FA April J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

FA Ballew P.M., Basu A., Baxendale J., Bayraktaroglu T., Beasley E.M.,

FA Benson K.V., Benos P.V., Bernier R.P., Bhandari D., Bolshakov S.,

FA Borke D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,

FA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

FA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

FA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

FA Dodson K., Doup L.E., Dowries M., Fagan-Pacha S., Dunkov B.C., Dunn P.,

FA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,

FA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

FA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

FA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,

FA Hoslin P., Houston K.A., Lowland T.J., Wei M.-H., Ikegawa C.,

FA Jafarli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

FA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

FA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

FA Liu X., Mattei H., McIntosh T.C., Melrod M.P., Melterson D.,

FA Merkulov G., Milshina N.V., Moharry C., Morris J., Moshrefi A.,

FA Mount S.M., Moy M., Murphy B., Murphy T., Muzny D.M., Nelson D.L.,

FA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,

FA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

FA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen R.,

FA Shue H.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

FA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

FA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

FA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

FA Williams S.M., Woldade T., Worley K.C., Wu D., Yang S., Yao Q.A.,

FA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF176816; AAF00990.1; .

KW Hypothetical protein; Coiled coil.

FT DOMAIN 47 1426 COILED COIL (POTENTIAL).

SO SEQUENCE 1453 AA; 16994 MW; 03CBA02A64CF4139 CRC64;

Query Match 0.5%; Score 10; DB 1; Length 1453;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 LQKIKKLES 742

DB 956 LQKIKKLES 955

|||||

RESULT 4

BP28_DROME STANDARD: PRI: 2096 AA.

AC Q9VM75;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein (CG10805).

GN CG10805.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Euryarchaea; Neoptera; Eulipeter; Diptera; Brachyera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

FX MEDLINE=20196006; PubMed=7071132;

FA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

FA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

FA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

FA Sutton G.G., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,

FA Brannon R.C., Rogers Y.-H.C., Blazej P.G., Champe M., Pfeiffer H.D.,

FA Wan K.H., Doyle C., Baxter E.G., Helt G., Helt G., Nikles G.L.G.,

FA April J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

FA Ballew P.M., Basu A., Baxendale J., Bayraktaroglu T., Beasley E.M.,

FA Benson K.V., Benos P.V., Bernier R.P., Bhandari D., Bolshakov S.,

FA Borke D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,

FA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

FA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

FA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

FA Dodson K., Doup L.E., Dowries M., Fagan-Pacha S., Dunkov B.C., Dunn P.,

FA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,

FA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

FA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

FA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,

FA Hoslin P., Houston K.A., Lowland T.J., Wei M.-H., Ikegawa C.,

FA Jafarli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

FA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

FA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

FA Liu X., Mattei H., McIntosh T.C., Melrod M.P., Melterson D.,

FA Merkulov G., Milshina N.V., Moharry C., Morris J., Moshrefi A.,

FA Mount S.M., Moy M., Murphy B., Murphy T., Muzny D.M., Nelson D.L.,

FA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,

FA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

FA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen R.,

FA Shue H.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

FA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

FA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

FA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

FA Williams S.M., Woldade T., Worley K.C., Wu D., Yang S., Yao Q.A.,

FA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,

Query Match 5.4%; Score 115; DB 1; Length 958;

Best Local Similarity 100.0%; Pred. No. 4.4e-107;

Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 QTHHALLLLCTVAGIPDKVILHNINSIFTEGANGVNRLLDTYSFQVINKTVKVPALJ 1358

DB 113 QTHHALLLLGTVAGIPDKVILHNINSIFTEGANGVNRLLDTYSFQVINKTVKVPALJ 172

|||||

QY 1359 QSDSPSPSEVSPNVEFTVKSISWVDALPHVDPHPPLPIIVGLVDTIGAEKFLW 1413

DB 173 QSDSPSPSEVSPNVEFTVKSISWVDALPHVDPHPPLPIIVGLVDTIGAEKFLW 227

|||||

RESULT 4

Y474_BOVIN STANDARD: PRI: 1453 AA.

AC Q9T023;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein KIAA0373.

GN KIAA0373.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Trachea;

FA Jovov B., Ripoll P.J., Benos E.J.;

FA Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF176816; AAF00990.1; .

KW Hypothetical protein; Coiled coil.

FT DOMAIN 47 1426 COILED COIL (POTENTIAL).

SO SEQUENCE 1453 AA; 16994 MW; 03CBA02A64CF4139 CRC64;

Query Match 0.5%; Score 10; DB 1; Length 1453;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 LQKIKKLES 742

DB 956 LQKIKKLES 955

|||||

RESULT 4

BP28_DROME STANDARD: PRI: 2096 AA.

AC Q9VM75;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein (CG10805).

GN CG10805.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Euryarchaea; Neoptera; Eulipeter; Diptera; Brachyera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

FX MEDLINE=20196006; PubMed=7071132;

FA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

FA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

FA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

FA Sutton G.G., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,

FA Brannon R.C., Rogers Y.-H.C., Blazej P.G., Champe M., Pfeiffer H.D.,

FA Wan K.H., Doyle C., Baxter E.G., Helt G., Helt G., Nikles G.L.G.,

FA April J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

FA Ballew P.M., Basu A., Baxendale J., Bayraktaroglu T., Beasley E.M.,

FA Benson K.V., Benos P.V., Bernier R.P., Bhandari D., Bolshakov S.,

FA Borke D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,

FA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

FA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

FA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

FA Dodson K., Doup L.E., Dowries M., Fagan-Pacha S., Dunkov B.C., Dunn P.,

FA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,

FA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

FA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

FA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,

FA Hoslin P., Houston K.A., Lowland T.J., Wei M.-H., Ikegawa C.,

FA Jafarli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

FA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

FA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

FA Liu X., Mattei H., McIntosh T.C., Melrod M.P., Melterson D.,

FA Merkulov G., Milshina N.V., Moharry C., Morris J., Moshrefi A.,

FA Mount S.M., Moy M., Murphy B., Murphy T., Muzny D.M., Nelson D.L.,

FA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,

FA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

FA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen R.,

FA Shue H.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

FA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

FA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

FA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

FA Williams S.M., Woldade T., Worley K.C., Wu D., Yang S., Yao Q.A.,

FA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,

RA Chen X H., Zhou F N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.: "
 RT The genome sequence of *Drosophila melanogaster*".
 CC Science 287:2185-2195(2000).
 CC -1- SIMILARITY: HELDINGS TO THE RAF28 FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HEAT REPEAT.
 CC -----
 CC This SWISS-PROT entry is copyright © It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF006b15; AAF52447.2; -;
 DR Flybase: FBgn0031864; C010805.
 DR InterPro: IPR001810; F-box.
 DR InterPro: IPR000357; HEAT_repeat.
 DR Pfam: PF00646; F-box; 1.
 DR PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.
 KW Hypothetical protein.
 FT REPEAT 2058 2094 HEAT
 FT REPEAT 2058 2094 HEAT
 SQ SEQUENCE 2095 AA: 237217 MW: 3E7B0657E54F520E P1: P64;

 Query Match: 0.4%, Score 9, DH 1, Length 2095;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
 QY 30 LRDPEPAT 38
 DB 31 LRDPEPAT 39

 RESULT 5
 CSF2_HUMAN
 ID CSF2_HUMAN STANDARD: PPT: 144 AA.
 AC P04141;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-NOV-1986 (Rel. 03, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Granulocyte-macrophage colony-stimulating factor precursor (GM-CSF)
 DE Granulocyte-stimulating factor (CSF) (Saragamostim) (Molgramostin).
 GN CSF2 OR GM-CSF.
 OS Homo sapiens (Human).
 CC Fukuyama, M.; Katayama, Y.; Choudhary, V.; Vertebata, Enteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE: 85242684; PubMed: 3925454;
 RA Lee F., Yokota T., Otsuka T., Gemmell L., Watson N., Loh J.,
 RA Arai K.-I., Rennick D.;
 RT "Isolation of cDNA for a human granulocyte-macrophage
 RT colony-stimulating factor by functional expression in mammalian
 RT cells".
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4360-4364(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86205844; PubMed: 3486413;
 RA Kaushansky K., O'Hara P. J., Berkner K., Segal G.M., Hagen F. S.,
 RA Adamson J.W.;
 RT "Genomic cloning, characterization, and multilineage growth-promoting
 RT activity of human granulocyte-macrophage colony-stimulating factor".
 PL Proc. Natl. Acad. Sci. U.S.A. 83:3101-3105(1986)
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85298429; PubMed: 3898382;
 RA Cantrell M.A., Anderson D., Cerretti D.P., Price V., McKereghan K.,
 RA Tushinski R.J., Mochizuki D.Y., Larson A., Grabstein K., Cosman D.;
 RT "Cloning, sequence, and expression of a human granulocyte/macrophage
 RT colony-stimulating factor".

FL Proc. Natl. Acad. Sci. U.S.A. 82:6250-6254(1985).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85218749; PubMed: 2923623;
 RA Wong C. G., Witte J. S., Temple P. A., Wilkens K. M., Leary A. C.,
 RA Luxenberg P. P., Jones S. S., Brown F. L., Kay P. M., Orr E. C.,
 RA Shemmaker C., Golde D. W., Kaufman P. J., Hewick P. M., Wang E. A.,
 RA Clark S. C.;
 RT "Human GM-CSF: molecular cloning of the complementary DNA and
 RT purification of the natural and recombinant proteins".
 RL Science 228:810-815(1985).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: A6030234; PubMed: 3876930;
 RA Miyatake S., Otsuka T., Yokota T., Lee F., Arai K.-I.;
 RT "Structure of the chromosomal gene for granulocyte macrophage colony
 RT stimulating factor: comparison of the mouse and human genes".
 RL EMBO J. 4:2561-2568(1985)
 RN [6]
 RP SEQUENCE FROM N.A.
 RX Kimmerly W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
 RA Davis C. A., Kadner K., Miguel T., Pittluck S., Ballard M., Wojeski H.,
 RA Subramanian S., Martin C.H.;
 PL Submitted (MAP 1998) to the EMBL/GenBank/DBS databases.
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE: 92144609; PubMed: 1737041;
 RA Kaushansky K., Lopez J.A., Brown C.B.;
 KI "Role of carbohydrate modification in the production and secretion of
 RT human granulocyte macrophage colony-stimulating factor in genetically
 RT engineered and normal mesenchymal cells".
 RL Biochemistry 31:1881-1886(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE: 92108429; PubMed: 1837174;
 RA Diederichs K., Boone T., Karplus P.A.;
 RT "Novel fold and putative receptor binding site of
 RT granulocyte-macrophage colony-stimulating factor".
 RL Science 254:1779-1782(1991).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE: 92235844; PubMed: 1569568;
 RA Waller M. P., Cook W. T., Fallick S. E., Nagabhushan T. L., Trotta P. P.,
 RA Bugg C. E.;
 RT "Three-dimensional structure of recombinant human granulocyte-
 RT macrophage colony-stimulating factor".
 RL J. Mol. Biol. 224:1075-1085(1992).
 CC -1- FUNCTION: CYTOKINE THAT STIMULATES THE GROWTH AND DIFFERENTIATION
 CC OF HEMATOPOIETIC PRECURSOR CELLS FROM VARIOUS LINEAGES, INCLUDING
 CC GRANULOCYTES, MACROPHAGES, EOSINOPHILS AND ERYTHROCYTES.
 CC -1- SUBUNIT: MONOMER.
 CC -1- POLYMORPHISM: VARIANT IL-3/IL-117 MAY BE A RISK FACTOR FOR APOPTIC
 CC ASTHMA.
 CC -1- PHARMACEUTICAL: Available under the names Leukine (Immunex) and
 CC Leucomax (Novartis) used in myeloid reconstitution following bone
 CC marrow transplant, bone marrow transplant engraftment failure or
 CC delay, mobilization and following transplantation of autologous
 CC peripheral blood progenitor cells, and following induction
 CC chemotherapy in older adults with acute myelogenous leukemia.
 CC -1- DATABASE: NAME-Leukine; NOTE-clinical information on Leukine;
 CC WWW="http://www.immunex.com/patient/pa02el.html".
 CC -----
 CC This SWISS-PROT entry is copyright © It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M13207; AAA98768.1; -;
 DR EMBL: M11734; AAA52122.1; -;
 DR EMBL: M11220; AAA52578.1; -;
 DP

DR EMBL: X04921; CAA26822.1;
 DR EMBL: M10663; AAA52121.1;
 DR EMBL: AC004511; AAC08707.1;
 DR PIR: A01854; ECHRM.
 DR PIR: G24646; G24646.
 DR PIR: A25169; A25169.
 DR PIR: L583; 31-JAN-94.
 DR PIR: G2MF; 08-NOV-96.
 DR MIM: 148960;
 DR InterPro: IPR000773; GM_CSF.
 DR Pfam: PF01109; GM_CSF; 1.
 DR PRINTS: PR00694; GM_CSF; 1.
 DR PROSITE: PS00749; GM_CSF; 1.
 DR SMART: SM00040; CSF2; 1.
 DR PROSITE: PS00702; GM_CSF; 1.
 KW cytokine; Growth factor; Glycoprotein. Signal: 3D-structure;
 KW Polymorphism: Pharmaceutical.
 FI SIGNAL 1 17
 FI CHAIN 18 144
 FT DISULFID 71 113
 FT DISULFID 105 134
 FT CARBOHYD 22 32
 FT CARBOHYD 24 24
 FT CARBOHYD 26 26
 FT CARBOHYD 27 27
 FT CARBOHYD 44 44
 FT CARBOHYD 54 54
 FT VARIANT 117 117
 FI TURN 25 27
 FI HELIX 40 41
 FI TURN 45 45
 FI HELIX 50 54
 FI STRAND 56 60
 FI HELIX 72 81
 FI TURN 82 82
 FI HELIX 85 103
 FI TURN 104 104
 FI STRAND 115 119
 FI HELIX 120 131
 FI TURN 132 133
 SO SEQUENCE 144 AA; 16295 MW; 75D1E50506CA7A8 CRC64;
 Query Match 0.4%; Score 8; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 8 1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1305 LLLGGTVA 1312
 Db 6 LLLGGTVA 13
 RESULT 6
 ROSE_CLOAR
 ID RPSE_CLOAR STANDARD: PPT: 245 AA
 AC P33657;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DI 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RNA polymerase sigma-E factor.
 GN SICE OR CAC1695.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OC NCBI_taxid=1488;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE-955189110; PubMed-7883192;
 RA Wong J., Sass C., Bennett G.N.;
 RT *Sequence and arrangement of genes encoding sigma factors in
 RT Clostridium acetobutylicum ATCC 824.*;
 RL Gene 153:89-92(1995).
 RN 12;
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RA Santangelo J.D., Kuhn A., Treuner A., Durte P.;
 RT *Sporulation and time course expression of sigma factor homologous
 RT genes in Clostridium acetobutylicum.*;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN 13;
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE-21359325; PubMed-11466286;
 RA Neill J., Reaugh G., Gmelchinsky M.V., Makarewicz R.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois T., Gu D., Hitti J., Wolf Y.L.;
 RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Sourcaille P., Daly M.J.,
 RA Benpett G.N., Keenan E.V., Smith D.P.;
 RT *Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.*;
 RL J. Bacteriol. 183:4823-4838(2001).
 RN 14;
 RP SEQUENCE OF 46-235 FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE-95050216; PubMed-7961408;
 RA Sauer G., Treuner A., Buchholz M., Santangelo J.D., Durte P.;
 RT *Sporulation and primary sigma factor homologous genes in Clostridium
 RT acetobutylicum.*;
 RL J. Bacteriol. 176:6572-6582(1994).
 CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED. THIS SIGMA FACTOR IS RESPONSIBLE FOR THE
 CC EXPRESSION OF SPOULATION SPECIFIC GENES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.sib.ch/announcements/>
 CC or send an email to license@sib.sib.ch).
 CC
 DR EMBL: U07420; AAC43309.1;
 DR EMBL: Z23079; CAA80617.1;
 DR EMBL: AE007679; AAK79661.1;
 DR PIR: S34304; S34309.
 DR InterPro: IPR006943; Sigma_70.
 DR Pfam: PF00140; sigma70; 1.
 DR PROSITE: PS00715; SIGMA70_1; 1.
 DR PROSITE: PS00716; SIGMA70_2; 1.
 KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
 KW DNA-binding; Sporulation; Complete proteome.
 FT DOMAIN 82 95 POLYMERASE CORE BINDING (P-1ENITAI).
 FT DNA_BIND 202 221 H-T-H MOTIF (HY SIMILARITY).
 FT CONFLICT 147 147 L -> P (IN REF. 4).
 SO SEQUENCE 235 AA; 26969 MW; C726E18E6C93A903 CRC64;
 Query Match 0.4%; Score 8; DB 1; Length 235;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 684 VEDLISVG 691
 Db 80 VEDLISVG 87
 RESULT 7
 YK67_CAEEL
 ID YK67_CAEEL STANDARD: PPT: 250 AA.
 AC P34345;
 DT 01-FEB-1994 (Rel. 28, Created)

use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M91598; AAC47504.1; 1;
 DR BSSP; P00560; 1QPG;
 DR InterPro; IPR001576; PGK;
 DR Pfam; PF00162; PGK; 1;
 DR PRINIS; PR00477; PHGLYCKINASE;
 DR PROSITE; PS00111; POLYGERATE_KINASE; 1;
 KW Transferase; Kinase; Glycolysis;
 SQ SEQUENCE 417 AA; 44772 MW; 3F308467FEC740BF CRC64;

Query Match 0.4%; Score 8; DB 1; Length 417;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1587 LLEKVNAL 1594
 |||||
 DB 226 LLEKVNAL 233

RESULT 10
 YV97_MYCPN STANDARD; PRT; 541 AA;
 AC 975695;
 DI 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Hypothetical lipoprotein MPN097 precursor (R02_or1541)
 GN MPN097 OR M0957;
 OS Mycoplasma pneumoniae;
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmataceae; Mycoplasma;
 CX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelfarb R., Hilbert H., Plagens H., Pirkle E., Li B.-C.,
 RT Hermann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";
 RI Nucleic Acids Res. 24:4420-4449(1996).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Potential).
 CC -!- SIMILARITY: BELONGS TO THE M3185 / M3260 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF000067; AA95795.1;
 KW Hypothetical protein; Lipoprotein; Membrane; Signal;
 FT SIGNAL 1 27 POTENTIAL;
 FT CHAIN 24 541 HYDROPHILIC LIPOPROTEIN MPN097;
 FT LIPID 23 24 N-ACYL-DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 541 AA; 53153 MW; AD2FCH2H820H1E9C CRC64;

Query Match 0.4%; Score 8; DB 1; Length 541;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1762 AALQKVVE 1769
 |||||

DB 50 AALQKVVE 57

RESULT 11
 YV97_MYCPN STANDARD; PRT; 563 AA;
 AC P71067;
 DI 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative L-lactate permease YV97;
 GN YV97;
 OS Bacillus subtilis;
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus;
 CX NCBI_TaxID=1423;
 RI [1]
 RP SEQUENCE FROM N.A.
 RA Denizot F.C.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 51-563 FROM N.A.
 RC STRAIN=168;
 RA Fabret C., Quentin Y., Chapel N., Guiseppe A., Haeche J., Denizot F.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

-!- FUNCTION: MAY PLAY A ROLE IN L-LACTATE TRANSPORT.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential)
 CC -!- SIMILARITY: BELONGS TO THE LIDP FAMILY OF TRANSPORTERS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Z94043; CAB08002.1;
 DR EMBL; Z71928; CA96486.1;
 DR EMBL; Z39121; CA91542.1;
 DR Subtilist; H011875; YV97;
 DR InterPro; IPR003804; Lactate_perm.
 DR Pfam; PF02652; Lactate_perm; 1;
 KW Hypothetical protein; Transport; Transmembrane; Complete proteome.

FT TRANSMEM 14 34 POTENTIAL;
 FT TRANSMEM 37 57 POTENTIAL;
 FT TRANSMEM 73 93 POTENTIAL;
 FT TRANSMEM 131 151 POTENTIAL;
 FT TRANSMEM 157 177 POTENTIAL;
 FT TRANSMEM 194 214 POTENTIAL;
 FT TRANSMEM 220 240 POTENTIAL;
 FT TRANSMEM 249 269 POTENTIAL;
 FT TRANSMEM 304 324 POTENTIAL;
 FT TRANSMEM 381 401 POTENTIAL;
 FT TRANSMEM 419 439 POTENTIAL;
 FT TRANSMEM 448 468 POTENTIAL;
 FT TRANSMEM 506 526 POTENTIAL;
 FT TRANSMEM 542 562 POTENTIAL;

SQ SEQUENCE 563 AA; 59761 MW; E4B0984059H19H08 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 563;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1757 LLSALAL 1764
 |||||
 DB 15 LLSALAL 22

RESULT 12
 CLP_TRYBB STANDARD; PRT; 868 AA;
 ID CLP_TRYBB

AC P31543;
 DT 01-JUL-1993 (Rel. 26, last sequence update)
 DT 01-JUL-1993 (Rel. 26, last sequence update)
 DT 15-JUL-1999 (Rel. 36, last annotation update)
 DE Heat shock protein 100 (CIP protein).
 GN HSP100.
 OS Trypanosoma brucei brucei.
 CC Eukaryota; Euklenozoa, Kinetoplastida, Trypanosomatidae; Trypanosoma.
 CC NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90239044; PubMed 2185473;
 PA Gomesan S., Squires C., Pichersky E., Carrington M., Hobbs M.,
 PA Matlick J.S., Dallymple B., Kuramitsu H., Shiroza T., Foster T.,
 PA Clark W.P., Ross B., Squires C.L., Maurizi M.R.;
 RT "Conservation of the regulatory subunit for the CIP ATP-dependent
 RT protease in prokaryotes and eukaryotes.";
 RL Proc Natl Acad Sci U S A 87:3513-3517(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-STIB 367H / J1TAR1,
 RX MEDLINE-98384019; PubMed-9719514;
 PA Redpath M.B., Carnall N., Webb H., Courel M., Amorim A., Guther M.L.,
 PA Cardoso de Almeida M.L., Carrington M.;
 RT "Conservation of genetic linkage between heat shock protein 100 and
 RT glycosylphosphatidylinositol-specific phospholipase C in Trypanosoma
 RT brucei and Trypanosoma cruzi.";
 RL Mol. Biochem. Parasitol. 94:113-121(1998).
 CC -1- SIMILARITY: BELONGS TO THE CIP/ATP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M92325; AAA30173.1;
 DR EMBL: AJ000080; CAA03906.1;
 DR PIR: R35905; E35905.
 DR InterPro: IPR003543; AAA
 DR InterPro: IPR003959; AAA.s.bfam.
 DR InterPro: IPR001270; CLP_AB.
 DR InterPro: IPR004176; CIP_N.
 DR Pfam: PF00004; AAA; 1.
 DR Pfam: PF02861; CIP_N; 2.
 DR PRINTS: PF06400; CLPFFQTEA-PA
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00870; CLPAB_1; 1.
 DR PROSITE: PS00871; CLPAB_2; 1.
 DR Chaperone, ATP binding, Repeat.
 FT DOMAIN 159 407 1.
 FT DOMAIN 529 720 1.
 FT NP_BIND 204 211 ATP (POTENTIAL).
 FT NP_BIND 603 610 ATP (POTENTIAL).
 SO SEQUENCE 868 AA; 96904 MW; 7E06D4AFE46E1881 CRC64;

 Query Match 0.4%, Score 8, DR 1, Length 868;
 Best Local Similarity 100.0%, Prod No. 30;
 Matches 8, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

 QY 827 HLLIGLFE 834
 D 11111111
 Db 109 HLLIGLFE 116

 RESULT 13
 HCYA_OCTDO
 ID HCYA_OCTDO STANJAKU, PRT, 1233 AA.
 AC P12659; Q25585;
 DT 01-OCT-1989 (Rel. 12, created)

DT 01-MAY-1991 (Rel. 18, last sequence update)
 DT 01-MAY-2002 (Rel. 41, last annotation update)
 DE Hemocyanin A-type, units ODE to ODC (Fragment).
 OS Octopus dofleini (Giant octopus).
 CC Eukaryota; Metazoa; Mollusca, Cephalopoda, Colepoda; Octopoda;
 CC Incirrata; Octopodidae; Octopus.
 CC NCBI_TaxID=6644;
 RN [1]
 RP SEQUENCE OF 1-834 FROM N.A.
 RX MEDLINE-91095437; PubMed-1898774;
 PA Lang W.H., van Holde K.E.;
 RT "Cloning and sequencing of octopus dofileini hemocyanin cDNA: derived
 RT sequences of functional units Ode and Odi.";
 RL Proc Natl Acad Sci U S A 88:244-248(1991).
 RN [2]
 RP SEQUENCE OF 835-1233 FROM N.A.
 RX MEDLINE-87088084; PubMed-3207675;
 PA Lang W.H.;
 RT "cDNA cloning of the Octopus dofileini hemocyanin: sequence of the
 RT carboxyl-terminal domain.";
 RL Biochemistry 27:7276-7282(1988).
 CC -1- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
 CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
 CC ARTHROPODS.
 CC -1- COFACTOR: CONTAINS TWO COPPER ATOMS PER FUNCTIONAL UNIT
 CC -1- SUBUNIT: DECAMERS OF LARGE IDENTICAL SUBUNITS (350 kDa), EACH
 CC CONTAINING 7 GLOBULAR OXYGEN-BINDING DOMAINS: GDA, GDB, GDC, ODD,
 CC ODE, ODF, AND ODG.
 CC -1- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M57288; AAA29391.1;
 DR EMBL: J02835; AAA29390.1;
 DR PIR: A31137; A31137.
 DR PIR: A28786; A28786.
 DR InterPro: IPR002227; Tyrosinase.
 DR Pfam: PF00264; Tyrosinase; 4.
 DR PROSITE: PS00497; TYROSINASE_1; 2.
 DR PROSITE: PS00498; TYROSINASE_2; 3.
 DR Glycylated Transferrin, Copper, Glycoprotein.
 KW Hemolymph; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 4 4 ODD
 FT DOMAIN 5 422 ODE
 FT DOMAIN 423 839 ODF
 FT DOMAIN 840 1233 ODG
 FT METAL 45 45 COPPER A (BY SIMILARITY).
 FT METAL 65 65 COPPER A (BY SIMILARITY).
 FT METAL 74 74 COPPER B (BY SIMILARITY).
 FT METAL 186 186 COPPER B (BY SIMILARITY).
 FT METAL 190 190 COPPER B (BY SIMILARITY).
 FT METAL 217 217 COPPER B (BY SIMILARITY).
 FT METAL 463 463 COPPER A (BY SIMILARITY).
 FT METAL 481 481 COPPER A (BY SIMILARITY).
 FT METAL 490 490 COPPER A (BY SIMILARITY).
 FT METAL 599 599 COPPER B (BY SIMILARITY).
 FT METAL 603 603 COPPER B (BY SIMILARITY).
 FT METAL 630 630 COPPER B (BY SIMILARITY).
 FT METAL 880 880 COPPER A (BY SIMILARITY).
 FT METAL 899 899 COPPER A (BY SIMILARITY).
 FT METAL 908 908 COPPER A (BY SIMILARITY).
 FT METAL 1008 1008 COPPER B (BY SIMILARITY).
 FT METAL 1012 1012 COPPER B (BY SIMILARITY).
 FT METAL 1039 1039 COPPER B (BY SIMILARITY).
 FT METAL 51 62 BY SIMILARITY.

CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
 CC OF VARYING REGULARITY AND ARE THOUGHT TO FORM A SINGLE-STRANDED
 CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
 CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
 CC DIFFERENT SPECIES.
 CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/licenses/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z18361; CAA79165.1; -
 CC EMBL: X51695; CAA35992.1; -
 CC PIR: A34209; A34209.
 CC PIR: S32633; S32633.
 CC PIR: A40691; A40691.
 CC HSSP: P02633; 3ICB.
 CC InterPro: IPR002048; EF-hand.
 CC InterPro: IPR001751; S100_CaBP.
 CC Pfam: PF00036; EF-hand_1.
 CC Pfam: PF01023; S100; 1.
 CC PROSITE: PS00018; EF-HAND; 1.
 CC PROSITE: PS00303; S100_CaBP; FALSE_NEG.
 CC Repeat: Calcium-binding.
 CC
 CC DOMAIN 1 91 S-100 LIKE.
 CC CA-HIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
 CC CA-HIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
 CC DOMAIN 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.
 CC REPEAT 413 448 1-1.
 CC REPEAT 449 476 1-2.
 CC REPEAT 477 504 1-3.
 CC REPEAT 505 532 1-4.
 CC REPEAT 533 560 1-5.
 CC REPEAT 561 588 1-6.
 CC REPEAT 589 616 1-7.
 CC REPEAT 617 644 1-8.
 CC REPEAT 645 678 1-9.
 CC REPEAT 679 706 1-10.
 CC REPEAT 707 742 1-11.
 CC REPEAT 743 771 1-12.
 CC REPEAT 772 796 1-13.
 CC REPEAT 797 832 1-14.
 CC DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.
 CC REPEAT 938 961 2-1.
 CC REPEAT 962 985 2-2.
 CC REPEAT 986 1021 2-3.
 CC REPEAT 1022 1044 2-4.
 CC REPEAT 1045 1067 2-5.
 CC REPEAT 1068 1090 2-6.
 CC REPEAT 1091 1121 2-7.
 CC REPEAT 1122 1144 2-8.
 CC REPEAT 1145 1167 2-9.
 CC REPEAT 1168 1197 2-10.
 CC REPEAT 1198 1227 2-11.
 CC REPEAT 1228 1250 2-12.
 CC REPEAT 1251 1273 2-13.
 CC REPEAT 1274 1296 2-14.
 CC REPEAT 1297 1319 2-15.
 CC REPEAT 1320 1342 2-16.
 CC REPEAT 1343 1368 2-17.
 CC REPEAT 1369 1391 2-18.
 CC REPEAT 1392 1416 2-19.
 CC REPEAT 1417 1439 2-20.

Query Match 0.4% Score 8; DB 1; Length 1233;

Best local Similarity 100.0%; Pref. No. 54;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 LEHLKEI 451

DB 1119 LEHLKEI 1126

RESULT 14

ID TRHY SHEEP

AC P22763; STANDARD; PRI: 1949 AA.

DI 01 AUG 1991 (REL. 19; Created)

DI 01 OCT 1994 (REL. 40; Last sequence update)

DI 16 OCT 2001 (REL. 40; Last annotation update)

DE Trichohyalin.

GN THH.

OS oviv aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Ciliartiodactyla; Kuminantia; Pecora; Bovoides;

OC Bovidae; Caprinae; oviv;

OX NR1_TAXID:9940;

RN [1]

RX MEDLINE-9326018; PubMed-7464041;

RA Fierz M.J., Metcalfe C.J., Campbell M.L., Rogers G.E.;

RT Analysis of the sheep trichohyalin gene: potential structural and

calcium-binding roles of trichohyalin in the hair follicle.;

RC J. Cell Biol. 121:855-865(1993).

RD [2]

RE SEQUENCE OF 1016-1549 FROM N.A.

RF STRAIN-MERINO-DORSET BORN X BORDER LEICESTER; TISSUE-WOOL follicles;

RX MEDLINE-9010682; PubMed-2298812;

RA Fierz M.J., Prestand R.B., Rogers G.E.;

RT The corna-derived amino acid sequence for trichohyalin, a

differentiation marker in the hair follicle, contains a 23 amino acid

repeat.;

R1 J. Cell Biol. 110:427-436(1990).

RC [1-]

CC FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES

IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE

INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR

LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY

ISOMERIZING BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, FOOTSTEP

WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN

ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN

ITS OWN CALCIUM DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL

DIFFERENTIATION.

CC 1 SUBUNIT: HOMODIMER (PROBABLE).;

CC 1 ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY

ALTERNATIVE SPLICING OF THE SAME GENE.

CC 1 TISSUE SPECIFICITY: FOUND IN THE HAD KAPPAINIZING TISSUES SUCH AS

THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN

THE EPITHELIA OF THE TONGUE, HOOF AND ROMEN.

CC 1 DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND

CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST

CC

```

FT REPEAT 1440 1461 2-21.
FT REPEAT 1462 1484 2-22.
FT REPEAT 1485 1507 2-23.
FT VARIANT 1145 1197 MISSING (IN SHORT FORM)
FT VARIANT 1251 1273 MISSING (IN SHORT FORM)
FT CONFLICT 1359 1399 K -> G (IN REF. 2)
SQ SEQUENCE 1549 AA, 20:173 MW, E72FB9FF1326E54E CRC64;

Query Match 0.4%, Score 8, DB 1; Length 1549;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1545 GLEERLLE 1552
DB 247 GLEERLLE 254
|||||||
|||||||

RESULT 15
HCYG_GCTDO STANDARD: PRT: 2896 AA.
ID HCYG_GCTDO STANDARD: PRT: 2896 AA.
AC 061363;
DT 30-MAY-2000 (Ref. 39, Created)
DI 01-MAR-2002 (Ref. 41, Last sequence update)
DE Hemocyanin G type, units ODA to ODC.
GN OBHCY.
OS Octopus dofleini (Giant octopus).
OC Eukaryota, Metazoa, Mollusca, Cephalopoda, Colepoda; Octopoda;
OC Invertebrata; Octopodidae; Octopus.
OX NCBI_TaxID=6644;
RN [1]
RP SEQUENCE FROM N. A. AND CHAPMAN-TERPITATION
RC TISSUE=Branchial gland;
RX MEDLINE=98277150; PubMed=9614945;
RA Miller K.E., Cuff M.E., Lang W.F., Varga-Weisz P., Field K.G.,
RA van Holde K.E.;
RT *Sequence of the Octopus dofleini hemocyanin subunit, structural and
RT evolutionary implications.*
RL J. Mol. Biol. 278:827-842(1998).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF SUBUNIT ODC.
RX MEDLINE=98277152; PubMed=9614947;
RA Cuff M.E., Miller K.E., van Holde K.G., Hendrickson W.A.;
RT *Crystal structure of a functional unit from octopus hemocyanin*;
CC ! FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
CC ARTHROPODS.
CC !- COFACTOR: CONTAINS TWO COPPER ATOMS PER FUNCTIONAL UNIT.
CC !- SUBUNIT: DIMERS OF LARGE IDENTICAL SUBUNITS (350 kDa). EACH
CC CONTAINING 7 GLOBULAR OXYGEN-BINDING FUNCTIONAL UNITS. ODA, ODB,
CC ODC, ODD, ODE, ODF, AND OEG; DEAMER FORMATION PEPTIDES THE
CC PRESENCE OF MAGNESIUM IONS.
CC !- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
CC SUBFAMILY.
CC
CC This SWISS-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF020548; AAC39018.1;
CC InterPro: IP000227, Tyrosinase.
CC Pfam: PF00264, Tyrosinase; 8.
CC PRINTS: PR00092; TYROSINASE.
CC PROSITE: PS00497, TYROSINASE_1; 4.
CC PROSITE: PS00498, TYROSINASE_2; 6.
CC Oxygen transport; Transport; Copper; Glycoprotein;
CC Hemolymph; Repeat.

```

```

PT DOMAIN 1 419
PT DOMAIN 420 834
PT DOMAIN 834 1254
PT DOMAIN 1255 1667
PT DOMAIN 1668 2085
PT DOMAIN 2086 2502
PT DOMAIN 2503 2896
PT METAL 41 60
PT METAL 60 69
PT METAL 69 178
PT METAL 178 182
PT METAL 182 209
PT METAL 209 460
PT METAL 460 480
PT METAL 480 489
PT METAL 489 601
PT METAL 601 605
PT METAL 605 632
PT METAL 632 875
PT METAL 875 895
PT METAL 895 904
PT METAL 904 1013
PT METAL 1013 1017
PT METAL 1017 1044
PT METAL 1044 1292
PT METAL 1292 1312
PT METAL 1312 1321
PT METAL 1321 1425
PT METAL 1425 1429
PT METAL 1429 1456
PT METAL 1456 1708
PT METAL 1708 1728
PT METAL 1728 1737
PT METAL 1737 1849
PT METAL 1849 1853
PT METAL 1853 1880
PT METAL 1880 2125
PT METAL 2125 2144
PT METAL 2144 2153
PT METAL 2153 2252
PT METAL 2252 2266
PT METAL 2266 2293
PT METAL 2293 2543
PT METAL 2543 2562
PT METAL 2562 2571
PT METAL 2571 2675
PT METAL 2675 2702
PT METAL 2702 57
PT DISULFID 47 168
PT DISULFID 168 323
PT DISULFID 323 466
PT DISULFID 466 591
PT DISULFID 591 881
PT DISULFID 881 1003
PT DISULFID 1003 1298
PT DISULFID 1298 1415
PT DISULFID 1415 1581
PT DISULFID 1581 1714
PT DISULFID 1714 1839
PT DISULFID 1839 1997
PT DISULFID 1997 2131
PT DISULFID 2131 2252
PT DISULFID 2252 2406
PT DISULFID 2406 2549
PT DISULFID 2549 2641
PT DISULFID 2641 2815
PT DISULFID 2815 60
PT THIOETH 60 478
PT THIOETH 478 895
PT THIOETH 895 1312
PT THIOETH 1312 1726
PT THIOETH 1726

```

ODA.

ODB.

ODC.

ODE.

ODE.

ODE.

ODG.

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER H (BY SIMILARITY).

COPPER H (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER A (BY SIMILARITY).

COPPER B (BY SIMILARITY).

COPPER B (BY SIMILARITY).

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

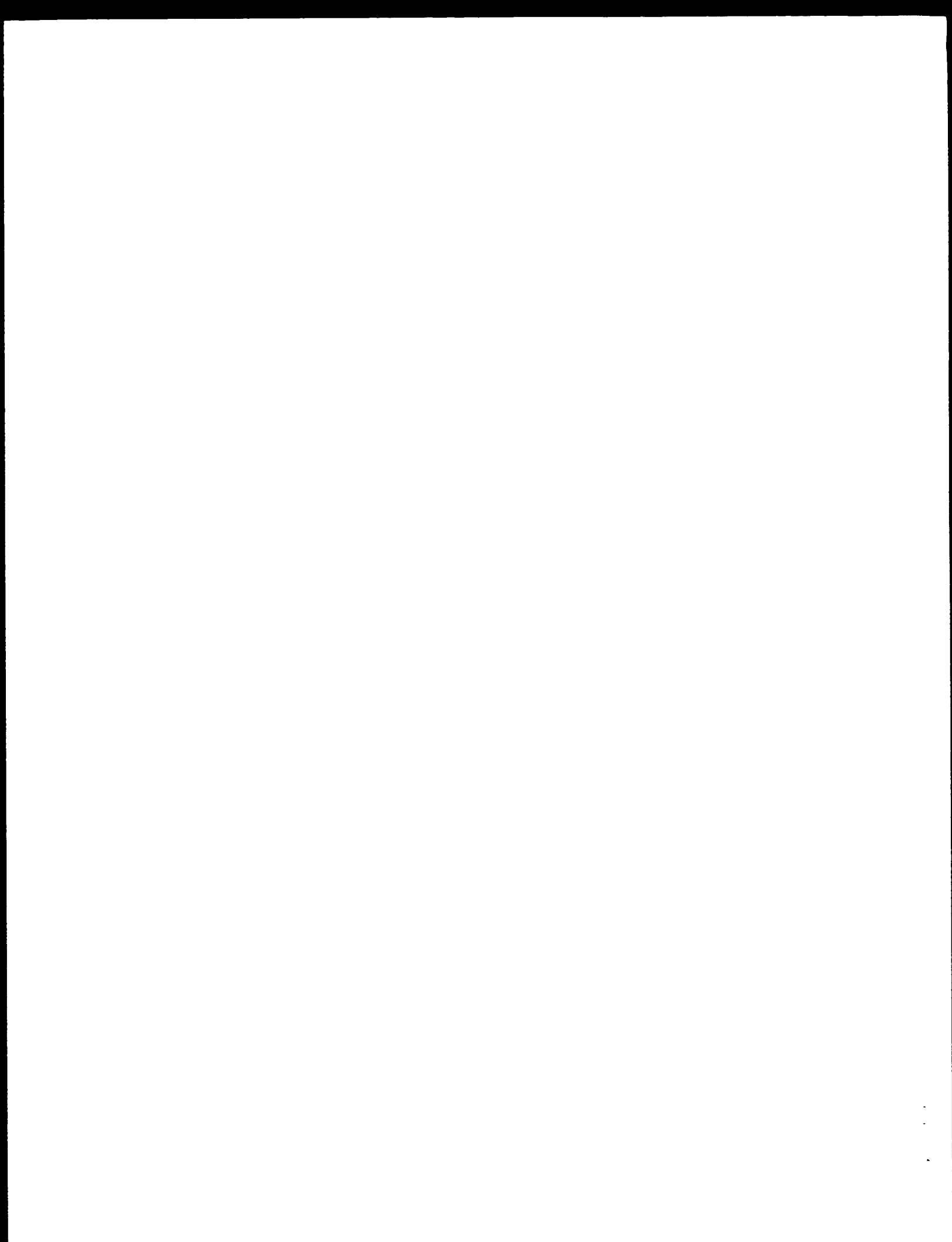
BY SIMILARITY.

FI THIOETH 2142 2144 BY SIMILARITY.
 FI THIOETH 2560 2562 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FI CARBOHYD 486 486 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FI CARBOHYD 804 804 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FI CARBOHYD 1496 1496 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FI CARBOHYD 1634 1634 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FI CARBOHYD 2055 2055 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FI CARBOHYD 2201 2201 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FI CARBOHYD 2553 2553 N-LINKED (GLCNAC: . .) (POTENTIAL).
 SQ SEQUENCE 2896 AA: 31917 MW: BELF45CHC987FBF3 GRC64;

Query Match 0.4%; Score 8; DH 1; Length 2896;
 Best Local Similarity 100.0%; Prod. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 LEEHLKEI 451
 DB 2782 LEEHLKEI 2789

Search completed: July 2, 2002, 15:42:58
 Job time: 375 sec



Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ENDPEKC 39

DB 29 ENDPEKC 35

RESULT 2

US-08-211-202-130
; Sequence 130, Application US/08211202
; Patent No. 5565332
; GENERAL INFORMATION:
; APPLICANT: HOOGENBOOM, Hendricus Renerus Jacobus Mattheus
; APPLICANT: HAIR, Michael
; APPLICANT: JESPER, Laurent Stephane Aure Therese
; APPLICANT: WINTER, Gregory Paul
; TITLE OF INVENTION: Production of chimeric antibodies - a
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David W Clough, Marshall C'Teale Gerstein Murray &
; ADDRESS: Bofun
; STREET: 6400 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/98/211,202
; FILING DATE: 23-SEP-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 9120252.3
; FILING DATE: 23-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9120377.8
; FILING DATE: 25-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 9206318.9
; FILING DATE: 24-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9206372.6
; FILING DATE: 24-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/00883
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: David W. Clough
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 2811/31960
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-211-202-130

Query Match 3.6%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred No 9 5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 KFOERVT 93

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

1 CITY: New York
2 STATE: NY
3 COUNTRY: USA
4 ZIP: 10174
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Diskette
7 OPERATING SYSTEM: DOS
8 SOFTWARE: FastSeq for Windows Version 2.0
9 CURRENT APPLICATION NUMBER: 60/092,415
10 APPLICATION NUMBER: US 09/347,843
11 FILING DATE: 12-SEP-1997
12 CLASSIFICATION: 435
13 ATTORNEY/AGENT INFORMATION:
14 NAME: Lambiris, Elias J.
15 REGISTRATION NUMBER: 44,728
16 REFERENCE/DOCKET NUMBER: 4944,200-US
17 TELECOMMUNICATION INFORMATION:
18 TELEPHONE: 212-867-0124
19 TELEFAX: 212-878-4655
20 INFORMATION FOR SEQ ID NO: 57:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 178 amino acids
23 TYPE: amino acid
24 STRANDEDNESS: Single
25 TOPOLOGY: linear
26 MOLECULE TYPE: No. 5958727e
27 US 09-928-692-57

Query Match 3.6% Score 7; DH 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 41 LLLQFL 47
DB 104 LLLQFL 110

RESULT 5
US 09 347 843 6
1 Sequence 6, Application US/09347843
2 Patent No. 6294658
3 GENERAL INFORMATION:
4 APPLICANT: Famodu, Layo O.
5 APPLICANT: Odell, Joan T.
6 TITLE OF INVENTION: Factors Involved in Gene Expression
7 FILE REFERENCE: BH-1172
8 CURRENT APPLICATION NUMBER: US/09/347,843
9 CURRENT FILING DATE: 1999-07-02
10 EARLIER APPLICATION NUMBER: 60/092,415
11 EARLIER FILING DATE: July 16, 1998
12 NUMBER OF SEQ ID NOS: 11
13 SOFTWARE: Microsoft Office 97
14 SEQ ID NO 6
15 LENGTH: 428
16 TYPE: PRT
17 ORGANISM: Glycine max
18 US 09 347 843 6

Query Match 3.6% Score 7; DH 4; Length 428;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 73 PLVDQLE 79
DB 357 PLVDQLE 363

RESULT 6
US 09 347 843 4
1 Sequence 4, Application US/09347843

1 Patent No. 6294658
2 GENERAL INFORMATION:
3 APPLICANT: Famodu, Layo O.
4 APPLICANT: Odell, Joan T.
5 TITLE OF INVENTION: Factors Involved in Gene Expression
6 FILE REFERENCE: BH-1172
7 CURRENT APPLICATION NUMBER: US/09/347,843
8 CURRENT FILING DATE: 1999-07-02
9 EARLIER APPLICATION NUMBER: 60/092,415
10 EARLIER FILING DATE: July 16, 1998
11 NUMBER OF SEQ ID NOS: 11
12 SOFTWARE: Microsoft Office 97
13 SEQ ID NO 4
14 LENGTH: 655
15 TYPE: PRT
16 ORGANISM: Oryza sativa
17 US 09-347-833-4

Query Match 3.6% Score 7; DH 4; Length 655;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 73 PLVDQLE 79
DB 583 PLVDQLE 589

RESULT 7
US-09-230-196-2
1 Sequence 2, Application US/09230196
2 Patent No. 6307035
3 GENERAL INFORMATION:
4 APPLICANT: Rauscher III, Frank J.
5 APPLICANT: Jensen, David E.
6 TITLE OF INVENTION: BRCA1 Associated Protein (BAP-1) and
7 TITLE OF INVENTION: Uses Therefor
8 NUMBER OF SEQUENCES: 47
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Howson and Howson
11 STREET: Spring House Corporate Cntr., Po Box 457
12 CITY: Spring House
13 STATE: Pennsylvania
14 COUNTRY: U.S.A.
15 ZIP: 19477
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: PatentIn Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/09/230,196
23 FILING DATE:
24 CLASSIFICATION:
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 60/022,997
27 FILING DATE: 02-AUG-1996
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 60/038,109
30 FILING DATE: 19-FEB-1997
31 ATTORNEY/AGENT INFORMATION:
32 NAME: Bak, Mary E.
33 REGISTRATION NUMBER: 31,215
34 PEPER-PLATTNER NUMBER: WS166BUSA
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: 215-540-9200
37 TELEFAX: 215-540-5818
38 INFORMATION FOR SEQ ID NO: 2:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 729 amino acids
41 TYPE: amino acid
42 TOPOLOGY: linear
43 MOLECULE TYPE: protein

us-09-603-665-5_copy_1950_2144.olig6.ra1

Wed Jul 3 09:02:34 2002

```
US-09-230-196-2
Query Match 3.6%; Score 7; DB 4; Length 729;
Best Local Similarity 100.0%; Prod No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LAEKLKE 148
Db 448 LAEKLKE 454
|||||

RESULT 8
US-08-928-692-59
; Sequence 59, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Deborah S.
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727, N. 5958727th Ave. of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944, 200-US
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 916 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727e
; US-08-928-692-58

Query Match 3.6%; Score 7; DB 2; Length 916;
Best Local Similarity 100.0%; Prod No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 LLLQFIL 47
Db 642 LLLQFIL 648
|||||

RESULT 10
US-07-596-467-6
; Sequence 6, Application US/07596467
; Patent No. 5306862
; GENERAL INFORMATION:
; APPLICANT: Chappell, J.
; APPLICANT: Saunders, Court A.
; APPLICANT: Cuellar, P.
; APPLICANT: Wolf, Fred P.
; TITLE OF INVENTION: Method and Composition for Increasing
; TITLE OF INVENTION: Sterol Accumulation in Higher Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Sulker & Milnamow
; STREET: 180 N. Stetson St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent ID Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/596,467
```

FILED DATE: 19901012
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Gansson, Edward P.
TELEPHONE: 312 616-5400
TELEFAX: 312 616-5460
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US 07 596 467-6

Query Match 3.6% Score 7: DB 1: Length 1045;
Best Local Similarity 100.0%, Pred. No. 84;
Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 8 HLKPPA 14
|||||
DB 10 HLKPPA 16

RESULT 11
US 07 944 474-6
Sequence 6, Application US/07944474
Patent No. 5449126
GENERAL INFORMATION:
APPLICANT: Chappell, J.
APPLICANT: Saunders, Court A.
APPLICANT: Cuellar, R.
TITLE OF INVENTION: Process and Composition for Increasing
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dressler, Goldsmith, Shore, Suter & Milnamow
STREET: 180 N. State St
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/944,474
FILING DATE: 19920814
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Gansson, Edward P.
REGISTRATION NUMBER: 29,381
TELEPHONE: 312 616 5400
TELEFAX: 312 616-5460
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US 07 944 474-6

Query Match 3.6% Score 7: DB 1: Length 1045;
Best Local Similarity 100.0%, Pred. No. 84;
Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 8 HLKPPA 14

DB 10 HLKPPA 16
|||||

RESULT 12
US-07-783-861C-6
Sequence 6, Application US/07783861C
Patent No. 5460949
GENERAL INFORMATION:
APPLICANT: Saunders, Court A.
APPLICANT: Woll, Fred R.
APPLICANT: Mukharji, Indrani
TITLE OF INVENTION: A Method and Composition for Increasing
TITLE OF INVENTION: the Accumulation of Squalene and Specific Sterols in
TITLE OF INVENTION: Yeast
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept.
STREET: 200 East Randolph St.
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60680-0703

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/783,861C
FILING DATE: 19911028
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,380
FILING DATE: 15-Nov-1990
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5460949vall H.
TELEPHONE: 312 856-718C
TELEFAX: 312 856-4972
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-861C-6

Query Match 3.6% Score 7: DB 1: Length 1045;
Best Local Similarity 100.0%, Pred. No. 84;
Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 8 HLKPPA 14
|||||
DB 10 HLKPPA 16

RESULT 13
US-07-928-930A-3
Sequence 3, Application US/07928930A
Patent No. 5344822

GENERAL INFORMATION:
APPLICANT: Levine, Daniel M.; Parker, Thomas S.;
APPLICANT: Rubin, Albert L.
TITLE OF INVENTION: Methods Useful in Endotoxin
TITLE OF INVENTION: Prophylaxis and Therapy
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York

CGNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/928,930A
FILING DATE: 19920812
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5344822man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: R000 206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-928-930A-3

Query Match 3.1%, Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 AEKLE 148
Db 10 AEKLE 15
|||||

RESULT 14
US-08-288-568-3
Sequence 3, Application US/08/288568
Patent No. 5506218
GENERAL INFORMATION:
APPLICANT: Parker, Thomas S.; Levine, Daniel M.;
APPLICANT: Rubin, Albert I.; Gordon, Bruce R.; Saal, Stuart D.
TITLE OF INVENTION: Methods useful in Endotoxin Based
TITLE OF INVENTION: Prophylaxis and Therapy
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb
MEDIUM TYPE: storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,568
FILING DATE: 10-AUGUST-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07453
FILING DATE: 9-AUGUST-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/928,930
FILING DATE: 12-AUGUST-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5506218man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: R000 211
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-288-568-3

Query Match 3.1%, Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 AEKLE 148
Db 10 AEKLE 15
|||||

TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-288-568-3

Query Match 3.1%, Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 AEKLE 148
Db 10 AEKLE 15
|||||

RESULT 15
US-08-487-461-3
Sequence 3, Application US/08487461
Patent No. 5587366
GENERAL INFORMATION:
APPLICANT: Parker, Thomas S.; Levine, Daniel M.;
APPLICANT: Rubin, Albert I.; Gordon, Bruce R.; Saal, Stuart D.
TITLE OF INVENTION: Methods Useful in Endotoxin Based
TITLE OF INVENTION: Prophylaxis and Therapy
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb
MEDIUM TYPE: storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,461
FILING DATE: June 7, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07453
FILING DATE: 9-AUGUST-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/928,930
FILING DATE: 12-AUGUST-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587366man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: R000 211.2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-487-461-3

Query Match 3.1%, Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 AEKLE 148
Db 10 AEKLE 15
|||||

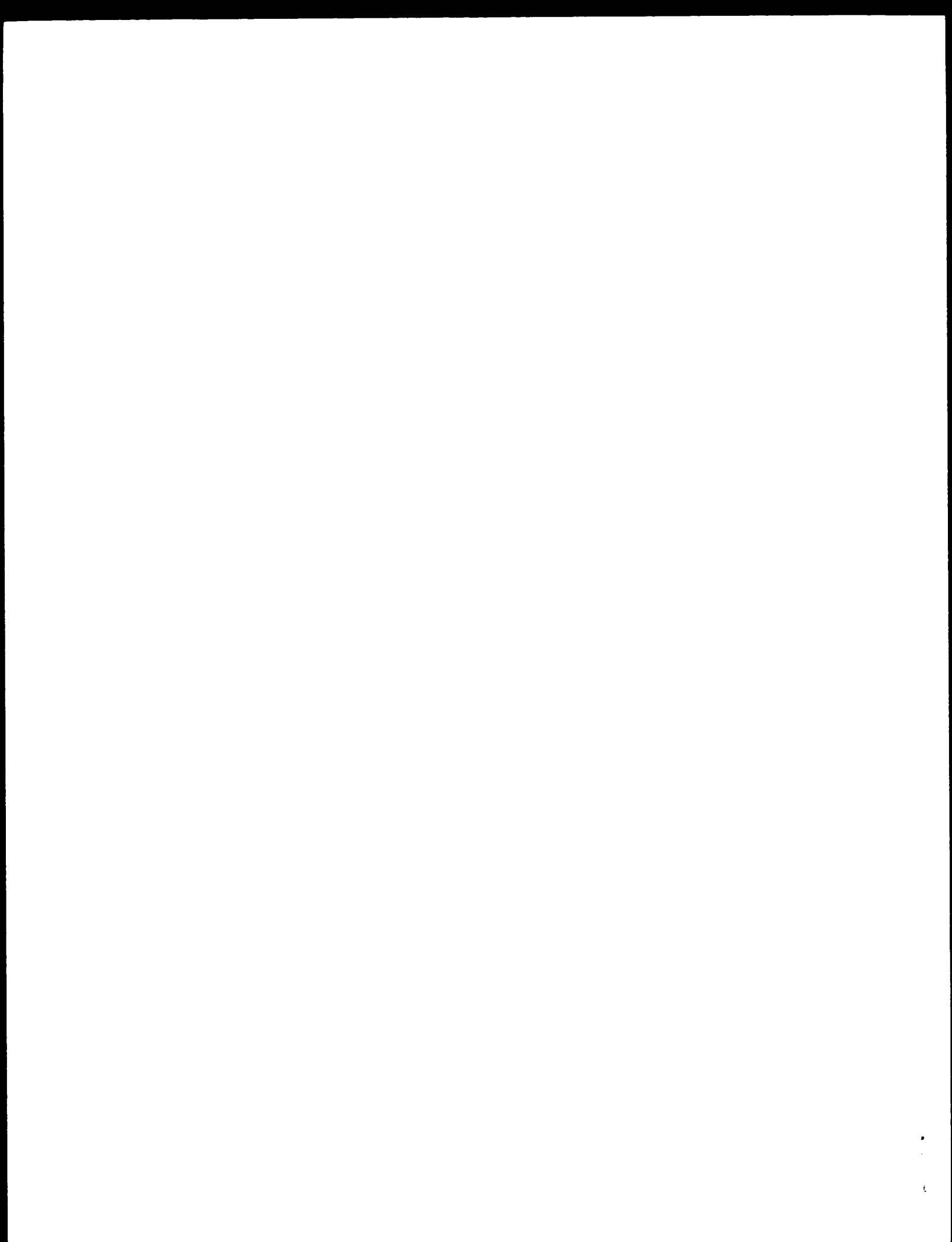
Wed Jul 3 09:02:34 2002

us-09-603-665-5_copy_1950_2144.olig6.rai

Page 7

14 10 AEKLE 15

Search completed: July 2, 2002, 16:04:02
Job time: 59 sec



Genome version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

MM protein - protein search, using sw model

Run on: July 2, 2002, 16:06:34 : Search time 117.69 seconds
(without alignments)
189,862 Million cell updates/sec

Files: US-09-603-665-5_COPY_1867_2067

Perfect score: 201

Sequence: 1 LIAFFLEALDFRAHRENDL.....CIAQSVAMALQSLWKPLNY 201

Scoring table:

Gapop 60.0 : Gapext 60.0

Searched: 747574 sets, 11074796 residues

Word size: 6

Total number of hits satisfying chosen parameters: 481

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database: A_Geneset_042802:*

```

1: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1940.DAT.*
2: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1941.DAT.*
3: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1942.DAT.*
4: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1943.DAT.*
5: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1944.DAT.*
6: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1945.DAT.*
7: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1946.DAT.*
8: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1947.DAT.*
9: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1948.DAT.*
10: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1949.DAT.*
11: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1950.DAT.*
12: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1951.DAT.*
13: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1952.DAT.*
14: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1953.DAT.*
15: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1954.DAT.*
16: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1955.DAT.*
17: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1956.DAT.*
18: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1957.DAT.*
19: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1958.DAT.*
20: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1959.DAT.*
21: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1960.DAT.*
22: /SDSL/qqdata/hold-geneseq/geneseq-emb1/AA1961.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	199	99.0	214	22	AAH85029
2	100	49.8	349	22	AAH92729
3	100	49.8	515	19	AAH54099
4	8	4.0	220	20	AAH24793
5	7	4.5	36	22	AAH40457
6	7	4.5	36	22	AAH61299
7	7	4.5	36	22	AAH74002
8	7	4.5	36	22	AAH34177
9	7	4.5	89	14	AAH34327
10	7	4.5	17	16	AAH66345
11	7	4.5	165	21	AAH27414

12	7	3.5	178	19	AAH37995
13	7	3.5	178	20	AAH39876
14	7	3.5	245	22	AAH95232
15	7	3.5	254	22	AAH35868
16	7	3.5	320	22	AAH17195
17	7	3.5	323	22	AAH24892
18	7	3.5	321	22	AAH27639
19	7	3.5	441	22	AAH92979
20	7	3.5	463	22	AAH39688
21	7	3.5	463	22	AAH66501
22	7	3.5	463	22	AAH93393
23	7	3.5	463	22	AAH70163
24	7	3.5	485	22	AAH41474
25	7	3.5	521	22	AAH9637
26	7	3.5	604	21	AAH27424
27	7	3.5	655	22	AAH13001
28	7	3.5	674	21	AAH42482
29	7	3.5	674	21	AAH03758
30	7	3.5	1045	15	AAH58611
31	6	3.0	18	40	AAH42447
32	6	3.0	28	21	AAH76326
33	6	3.0	32	22	AAH15816
34	6	3.0	45	22	AAH31446
35	6	3.0	51	22	AAH44072
36	6	3.0	51	22	AAH21705
37	6	3.0	53	22	AAH5086
38	6	3.0	57	21	AAH13000
39	6	3.0	70	22	AAH90952
40	6	3.0	71	22	AAH04594
41	6	3.0	79	21	AAH22312
42	6	3.0	82	22	AAH92148
43	6	3.0	85	21	AAH2086
44	6	3.0	85	22	AAH92711
45	6	3.0	86	22	AAH07992

ALIGNMENTS

RESULT 1
AAH85029
ID AAH85029 standard; Protein: 2144 AA.
XX
AAH85029;
XX
06-AUG-2001 (first entry)
XX
Protein encoded by BAP28 cDNA consisting of exons 1 to 45.
XX
BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.
XX
Homo sapiens.
XX
Key Location/Qualifiers
FT Misc-difference 1694
FT /label- Ser or Asn
FT Misc-difference 1854
FT /label- Ala or Val
FT Misc-difference 1967
FT /label- Asp or Asn
FT Misc-difference 2017
FT /label- Gly or Glu
XX
W0200100669-A2.
XX
04-JAN-2001.
XX
23 JUN 2000. 2603W0-IB01183.
XX
25-JUN-1999; 990S-0141423.
XX
18-JAN-2000; 2000US-0176R80.
XX
(GIST) GENSET.

Mutant Aspergillus
A. oryzae p7-14.1
Human protein sequ
Helicobacter pylori
Peptide #3629 enco
Peptide #3674 enco
Novel human diagno
Human protein sequ
Human polypeptide
App specific suve1
Human protein sequ
DNA encoding human
Human polypeptide
Cerephila melano
H. vulgaris NIM1 h
Pice poly (A) band
Human BFX 0042246
Human BFX 0042246
Human BFX 0042246
Yeast BMS-coA redu
Class A amphipath
Fragment of human
Human testous syst
Novel human secret
Peptide #11578 en
Peptide #8139 enco
Human immunoglob
Arabidopsis thaliana
Human immunoglob
Human polypeptide
Zea mays protein 1
Human digestive sy
Zea mays protein 1
Novel human diagno
Human polypeptide

XX Barry C, Bouquelieret L, Chumakov I, Cohen-Akenline A;
 PI WPI: 2001-367032/38.
 DR N-PSDB; AAF84909, AAF83910.
 XX New BAP28 polynucleotides and polypeptides overexpressed in prostate
 PT cancer cells for diagnosing prostate tumors, e.g. by hybridization or
 PT polymerase chain reaction assays -
 XX Claim 14; page 297-304; 349pp; English.
 PS The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
 CC sequences and regulatory region located at the 3' and 5' ends of the
 CC BAP28 coding region, the BAP28 polypeptides can be expressed by standard
 CC recombinant methodology, BAP28 polynucleotides and polypeptides have been
 CC found to be over expressed in prostate tumour cells, therefore levels of
 CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
 CC reaction (PCR)) to diagnose patient suffering from or susceptible to
 CC prostate cancer. Antibodies specific for the BAP28 polypeptides are
 CC useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
 CC useful in genetic analysis, the present sequence represents a protein
 CC encoded by a first cDNA sequence of the BAP28 gene consisting of the
 CC exons 1 to 45.
 XX Sequence 2144 AA:
 SQ
 Query Match 99.9%; Score 199; DB 22; Length 2144;
 Best Local Similarity 100.0%; Pred No 4.7e-197;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 UTAFFPLEALDFFRAQHSNDLEEVGKTENCIDCLVAMVVKISKVTFRPLFFKLFDAWAKTE 60
 DB 1867 tafflealdffraqhsndleevgktencidclvamvkvksevtfrplffkldwakte 1926
 QY 61 DAPKDPLETFYNIADCTAEKLGKLTFLFAGSHLVKPPADTLXQVNISSKTDCAFFDSNDPE 120
 DB 1927 dapkdprlctfynladctaecklgkltflfaghivkpfadtlxqvnissktdcaffdsndpe 1986
 QY 121 KCCLLIQTLNCLYKIFPTDHPISKEPAXALMMPILVDLENPLDGGPKPQRRVTKHLI 180
 DB 1987 kccillqtlnclykifptdhpiskpaxalmmpilvdlenpldggpkpqrvtkhli 2046
 QY 181 KCLAPSVAMADSLWKPLNY 201
 DB 2047 pclagfsvamadslwkpiny 2067
 RESULT 2
 AAH92729
 ID AAH92729 standard; protein; 349 AA.
 XX
 AC AAB92729;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:11159.
 XX
 KW Human; primer: detection; diagnosis; antitense therapy; gene therapy
 XX
 OS Homo sapiens
 XX
 PN KP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PE
 XX 29-JUL-1999; 99JP-0248036.
 PR
 XX 27-AUG-1999; 99JP-0300253.
 PR
 XX 11-JAN-2000; 2000JP-0118776.
 PR
 XX 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI: 2001-318749/34.
 XX primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX Claim 8; SEQ ID 11159; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification, the primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 349 AA:
 SQ
 Query Match 49.8%; Score 100; DB 22; Length 349;
 Best Local Similarity 100.0%; Pred No 3.6e-95;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 UTAFFPLEALDFFRAQHSNDLEEVGKTENCIDCLVAMVVKISKVTFRPLFFKLFDAWAKTE 60
 DB 72 tafflealdffraqhsndleevgktencidclvamvkvksevtfrplffkldwakte 131
 QY 61 DAPKDPLETFYNIADCTAEKLGKLTFLFAGSHLVKPPADTL 100
 DB 132 dapkdprlctfynladctaecklgkltflfaghivkpfadtl 171
 RESULT 3
 AAH54099
 ID AAH54099 standard; protein; 515 AA.
 XX
 AC AAW54099;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE Homo sapiens BAP28 sequence.
 XX
 KW BAP28; ring protein; RING1; breast cancer; risk; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN W09812327-A2.
 XX
 PD 26-MAR-1998
 XX

PE 19-SEP-1997: 97W0-0516842.
 XX
 PR 04-APR-1997: 97W0-0042985.
 PR 20-SEP-1996: 96W0-0026246.
 PR 03-APR-1997: 97W0-0042611.
 XX
 (TEXA) UNIV TEXAS SYSTEM.
 PA
 PI Baer R, Bowcock AM;
 XX
 DR WPI: 1998-230317/20.
 DR N-PSDB: AAY24115.
 XX
 PT DNA sequence encoding HARD1, B123, BE2, BE14, BE31 or BE445 - which
 PT as breast cancer antigen. BRCAL binding proteins are useful to
 PT identify patient having or at risk of developing cancer.
 XX
 PS Disclosure: Page 287-288; 34pp; English.
 XX
 CC The sequence is that of a protein which can be used in the
 CC preparation of the recombinant breast cancer antigen, BRCAL, binding
 CC proteins HARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the
 CC detection of a BARD1, B123, BE2, BE14, BE31 or BE445 nuclear acid
 CC sequence, specifically a wild type BARD1 composition for the detection
 CC or purification of BRCAL, useful to identify a patient having, or at
 CC risk of developing cancer. BARD1 can be used in the preparation of an
 CC anti-BARD1 antibody, and in the detection and purification of a BRCAL
 CC protein. BARD1, B123, BE2, BE14, BE31 or BE445 can be used in the
 CC identification of a binding protein agonist or antagonist that alters
 CC the binding of HARD1, B123, BE2, BE14, BE31 or BE445 to BRCAL or the
 CC biological activity of the BRCAL-HARD1, B123, BE2, BE14, BE31 or BE445
 CC complex. The antibodies can be used to detect BARD1, B123, BE2, BE14,
 CC BE31 or BE445, a specific anti-BARD1 antibody can be used to identify
 CC a patient having or at risk of developing cancer.
 XX
 SO Sequence 515 AA;

Query Match 49.8%; Score 100; DR 19; Length 515;
 Best Local Similarity 100.0%; Pred. No. 5, 2e-95;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LIAFFLEFALPRAHSENDLEEVCKIPNCLIDCLVAMVVKLSVTFPPLFKIFDWAKTE 60
 DB 248 LIAFFLEFALPRAHSENDLEEVCKIPNCLIDCLVAMVVKLSVTFPPLFKIFDWAKTE 60
 QY 61 DAPKQGLITFYNLAEITAEKLGHTFAGHLVKKPEADTL 100
 DB 298 DAPKQGLITFYNLAEITAEKLGHTFAGHLVKKPEADTL 100

RESULT 4
 AAY24115
 ID AAY24115 standard; Protein: 220 AA.
 XX
 AC AAY24115;
 XX
 DE 14-SEP-1999 (first entry)
 DE A gyrase protein of Chitinophaga pinensis.
 XX
 KW Identification; detection; microbe; gyrase gene; gyrase protein
 XX
 OS Chitinophaga pinensis.
 XX
 PN J011169175.A.
 XX
 PD 29-JUN-1999.
 XX
 PE 12-DEC-1997: 97JP 0143416.
 PR 12-DEC-1997: 97JP 0143416.
 XX

PA (MARU-) MARINE BIOTECHNOLOGY INST CO LTD.
 PA (KALY-) KALY BIOTECHNOLOGY KENKYUSHO KK.
 XX
 DE WPI: 1999-422615/36.
 DR N-PSDB: AAX86000.
 XX
 PT Identification and detection of a microbe - by detection of a
 PT gyrase gene
 XX
 PS Example 3: Page 14-15; 42pp; Japanese.
 XX
 CC The specification describes a method for the identification or
 CC detection of a microbe, using the gyrase gene as the index. The
 CC method involves the use of PCR primers to amplify DNA from the
 CC microbe, which is then identified or detected depending on its
 CC base sequence. The method can be used to classify and identify
 CC an unidentified microbe strain rapidly and with high precision.
 CC The present sequence represents a gyrase protein.
 XX
 SO Sequence 220 AA;

Query Match 4.0%; Score 8; DR 20; Length 220;
 Best Local Similarity 100.0%; Pred. No. 5, 4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SENDPEKC 122
 DB 130 sendpekc 137

RESULT 5
 ABB40457
 ID ABB40457 standard; Peptide: 36 AA.
 XX

AC ABB40457;
 XX
 DE 04-FEB-2002 (first entry)
 DE Peptide #7963 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN W0200157277-A2.
 XX
 PD 09-AUG-2001.
 XX

PE 30-JAN-2001: 2001W0-US00669.
 XX
 PR 04-FEB-2000: 2000US-0180312.
 PR 26-MAY-2000: 2000US-0207476.
 PR 30-JUN-2000: 2000US-0608408.
 PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000US-0024243.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX

PI Penn SG, Hanzel DK, Cher W, Rank DR;
 XX
 DR WPI: 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver.
 XX
 PS Claim 27: SEQ ID NO 33092: 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at http://wipo.int/pub/published_pct_sequences.

XX Sequence 36 AA;

Query Match 3.5%; Score 7; DB 22; Length 36;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGEKFPQ 172
 DB 4 ggeckfq 10
 |||||

RESULT 6
 AAM61279
 ID AAM61279 standard; Protein: 36 AA.
 AC AAM61279;
 XX
 XX
 XX 05-NOV-2001 (first entry)
 DT
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33384.

XX Human, brain expressed exon, gene expression analysis, probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.

XX Homo sapiens.
 OS
 PN WO200157275-A2.
 XX
 XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0604008.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DK;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4; SEQ ID NO: 33384, 650pp; Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is a protein encoded by one of

XX the probes of the invention.

XX Sequence 36 AA;

Query Match 3.5%; Score 7; DB 22; Length 36;
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGEKFPQ 172
 DB 4 ggeckfq 10
 |||||

RESULT 7
 AAM74002
 ID AAM74002 standard; Protein: 36 AA.

XX AAM74002;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 34308.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0604008.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 34308, 658pp; Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX bone marrow. They can be used to measure gene expression in bone marrow

XX samples, which may enable the improved diagnosis and treatment of cancers

XX such as lymphoma, leukaemia and myeloma. The present sequence is a

XX protein encoded by one of the probes of the invention.

XX Sequence 36 AA;

Query Match 3.5%; Score 7; DB 22; Length 36;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGEKFPQ 172
 DB 4 ggeckfq 10
 |||||

RESULT 8
 AAM34177
 ID AAM34177 standard; Protein: 36 AA.

XX AAM34177;

XX 17-OCT-2001 (first entry)

XX Peptide #214 encoded by probe for measuring placental gene expression.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.

XX Homo sapiens.
 XX W0200157272.A2.
 XX PD 09 AUG-2001.
 XX
 XX 10 JAN-2001; 2001WO 0500663.
 XX
 XX 04 FEB-2000; 2000RS 0180412.
 XX 26 MAY-2000; 2000RS-0207455.
 XX 10 JUN-2000; 2000RS-0608408.
 XX 04 AUG-2000; 2000RS-0632366.
 XX 21 SEP-2000; 2000RS-0246487.
 XX 27 SEP-2000; 2000RS-0246454.
 XX 04 OCT-2000; 2000GB 0024263.
 XX
 XX (MOLE) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001 408897/53.
 XX
 XX Human genome derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta .
 XX
 XX Claim 27: SEQ ID NO 34445; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs-
 XX see AA141315-AA157546). The present sequence is a peptide encoded by one
 XX such probe. The probes are useful for producing a microarray for
 XX predicting, measuring and displaying gene expression in samples derived
 XX from human placenta. The probes are useful for antenatal diagnosis of
 XX human genetic disorders.

XX Sequence 46 AA;
 XX
 XX Query Match 3.5%; Score 7; DB 22; Length 36;
 XX Best Local Similarity 100.0%; Pred. No. 11;
 XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GDEKFO 172
 DB |||||
 DB 4 qdektq 10

RESULT 9
 AAR44271
 ID AAR44271 standard; Protein: 89 AA.
 AC
 XX AAR44271;
 XX
 XX 26 JUL-1994 (first entry)
 XX
 XX bp 74 VH1 gene heavy chain.
 XX Human monoclonal antibody; anti globulin response;
 KW chimeric; mouse human antibodies; antibody; prevention
 KW
 XX Homo sapiens.
 XX
 XX W0906214 A.
 XX
 XX 01 APR-1994.
 XX
 XX 23 SEP-1992; 92WO-GB01765.
 XX

PR 23-SEP-1991; 91OR-0020252.
 PR 25-SEP-1991; 91GB-0020377.
 PR 24-MAR-1992; 92GB-0006319.
 PR 24-MAR-1992; 92GB-0006372.
 PR 15-MAY-1992; 92WO-GB00883.
 XX
 XX (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 XX (MEDI-) MEDICAL RES COUNCIL.
 XX
 XX Balier M, Hoozenboom HRJM, Jaspers LSAT, Winter GP;
 XX WPI: 1993-117534/14.
 XX
 XX Producing human antibody polypeptide dimer specific for antigen
 XX comprises use of chain shuffling using phage expression, useful
 XX for reducing anti globulin responses in humans for increased
 XX human characteristics
 XX
 XX Example: Fig 11; 109pp; English.
 XX
 XX The sequence is that of the DP-74 VH1 gene heavy chain
 XX which may be used as part of a method of producing chimeric
 XX mouse-human antibodies or fragments which have the same
 XX binding specificity as a parent Ab but have increased human
 XX characteristics, preventing anti-globulin response in humans.
 XX
 XX Sequence 89 AA;

XX Query Match 3.5%; Score 7; DB 14; Length 89;
 XX Best Local Similarity 100.0%; Pred. No. 25;
 XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KPQEPVT 176
 DB |||||
 DB 63 kfqrvt 69

RESULT 10
 AAR66445
 ID AAR66445 standard; Protein: 117 AA.
 AC
 XX AAR66445;
 XX
 XX 04-AUG-1995 (first entry)
 XX
 XX Human immunoglobulin variable heavy chain #51.
 XX
 XX Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
 KW cosmid; placenta; vector; pJH81; E.coli; mammalian.
 KW
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Misc-difference 15 /note "encoded by ACA codon"
 XX
 XX W09426895-A.
 XX
 XX 24 NOV-1994.
 XX
 XX 10-MAY-1993; 93WO-JP00603.
 XX
 XX 10 MAY-1993; 93WO-JP00603.
 XX
 XX (NISR) JAPAN TOHACCO INC.
 XX
 XX Honjo T, Matsuda F;
 XX WPI: 1995-006791/01.
 XX N-PSDB; AAQ78996.
 XX
 XX DNA fragment comprising human immunoglobulin VH genes - for the

PT production of human immunoglobulin in mammalian hosts

XX

PS Disclosure; Page 105-107; 130pp; Japanese.

XX

CC protein sequences (AA66295-51) are novel human immunoglobulin heavy

CC chain sequences encoded by novel isolated genes. The genes

CC (AA078949-79002) were isolated and cloned from a series of cosmid

CC constructs; Y202; Y103; Y21; Y6; Y24; 3-31; M84; M118 and M131, by PCR

CC amplification using primers AA078917-38, the genes are subdivided into 5

CC families of Vh genes. The fragments cover a region of 800 kb. The DNA

CC fragments were isolated from high molecular weight DNA from human

CC placenta. The DNA was partially digested with TaqI restriction enzyme.

CC The fragments were separated by gel electrophoresis and 35-45 kb fractions

CC were collected. The fragments were ligated with ClaI-digested cosmid

CC vector pJH81. The ligation products were in vitro packed and infected

CC into E. coli 490A. The fragments were then subcloned by colony

CC hybridisation. The Vh genes and the DNA fragments encoding them are

CC useful in producing human immunoglobulin in mammalian hosts.

XX

SQ Sequence 117 AA;

Query Match 3.5%, Score 7, DB 16, Length 117,
Best Local Similarity 100.0%, Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 170 KPOFRVT 176

DB 82 Kfgcvt 88

|||||

RESULT 11

AA027313

ID AAB27313 standard; Protein; 165 AA.

XX

AC AAB27313;

XX

DT 25 JAN 2001 (first entry)

DE H. vulgaris NIM1 homologue SEQ ID NO: 40.

XX

XX Systemic acquired resistance; SAR; signal transduction cascade;

KW disease resistance, tobacco, tomato, canola; sunflower; sugarbeet;

KW potato.

XX

OS Beta vulgaris.

XX

PN W0200053762-A2.

XX

PD 14-SEP-2000

XX

PP 07-MAR-2000; 2000W0-0301978.

XX

PP 06-MAR-1994; 94NS-0265149

XX

XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-ERFINDUNCEN VERW GES MBH.

XX

PI Salmann JM, Weislo LJ, Willits MG, Mangiste I,

XX

XX WPI: 2000-594322/56.

DR N-PSDB; AAA97216.

XX

XX Novel plant genes for enhancing systemic acquired resistance gene

PT expression and broad spectrum disease resistance in plants, are

PT homologues of Arabidopsis NIM1 gene -

XX

XX claim 1; Page 111; 154pp; English.

PS

PS the present invention is concerned with the isolation of NIM1 homologues

CC and their coding sequences from Arabidopsis thaliana, Brassica napus,

CC Nicotiana tabacum, Lycopersicon esculentum, Beta vulgaris, Helianthus

CC annuus and Solanum tuberosum. NIM1 is one of the proteins involved in the

CC

CC signal transduction cascade leading to systemic acquired resistance (SAR)

CC in plants. This gives the plants an increased resistance to disease. The

CC protein and gene can be used to produce transgenic plants resistant to

CC diseases caused by viruses, such as tobacco or cucumber mosaic virus,

CC tobacco etch virus, pelargonium leaf curl virus, red clover mottle virus,

CC tomato bushy stunt virus, fungi, including Phytophthora parasitica and

CC peronospora tabacina, bacteria such as Pseudomonas syringae and P.

CC tabaci, insects, including aphids and lepidoptera and nematodes such as

CC Meloidiomyne incognita. In particular they can be used against disease

CC organisms of maize.

XX

SQ Sequence 165 AA;

Query Match 3.5%, Score 7, DB 21, Length 165;
Best Local Similarity 100.0%, Pred. No. 45,
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 DAPKRL 67

DB 115 dapkrl 121

|||||

RESULT 12

AAW37995

ID AAW37995 standard; Protein; 178 AA.

XX

AC AAW37995;

XX

DT 09-SEP-1998 (first entry)

DE Mutant Aspergillus oryzae P7-14 1 rescued locus.

XX

XX Rescued locus; mutant; metabolite.

OS Aspergillus oryzae.

XX

PN W09811203-A1.

XX

PD 19-MAR-1998.

XX

PP 12-SEP-1997; 97W0-051004

XX

PP 13-SEP-1996; 96US-0713312.

XX

PA (NOVO) NOVO NORDISK BIOTECH INC

XX

PI Brody H, Hansen K, Lamsa MH, Yaver DS;

XX

DR WPI: 1998-207372/18.

DR N-PSDB; AAV31842.

XX

XX Polypeptide production - using a mutant cell with a nucleic acid

PT constructer introduced into a locus

XX

XX Disclosure; Fig 32; 195pp; English.

PS

XX This is the amino acid sequence of a rescued locus from the mutant

CC Aspergillus oryzae P7 14.1, used in the method of the invention which

CC involves the use of a mutant cell with a nucleic acid construct

CC introduced into a locus. The methods can be used for producing

CC metabolites.

XX

SQ Sequence 178 AA;

Query Match 3.5%, Score 7, DB 19, Length 178,
Best Local Similarity 100.0%, Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 124 LLLQFIL 130

DB 104 Lllqfil 110

|||||

RESULT 14
 AAY 9876
 ID AAY 9876 standard; Protein: 178 AA.
 XX
 AC AAY 9876;
 XX
 DT 04-DEC-1999 (first entry)
 XX
 DE A. oryzae P7 14.1 locus protein sequence.
 XX
 KW Protein production; mutant protein; DNA locus.
 XX
 OS Aspergillus oryzae
 XX
 PN DS5958727 A.
 XX
 PU 28-SEP-1999.
 XX
 PF 12-SEP-1997; 97DS-028692.
 XX
 PR 13-SEP-1996; 96DS-0713312.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 XX
 PI Brody H., Lamsa M., Hansen K., Yaver DS;
 XX
 DR WPI: 1999-56104/47.
 DR N PSDB: AA220711.
 XX
 PS Modifying the production of a polypeptide -
 XX
 PS Disclosure: Fig 33; 129pp; English.
 XX
 CC This sequence is encoded by an Aspergillus oryzae DNA locus. The
 CC invention relates to a method of producing a polypeptide, comprising:
 CC (a) cultivating a mutant cell under conditions conducive for production
 CC of the polypeptide, where: (i) the mutant cell is related to a parent
 CC cell, which comprises a first DNA sequence encoding the polypeptide, by
 CC the introduction of a nucleic acid construct into the genome of the
 CC parent cell at a locus which is not within the first DNA sequence, and
 CC not within a second DNA which encodes a protein that negatively regulates
 CC transcription, translation or secretion of the polypeptide, and not
 CC within a third DNA sequence encoding a protease which hydrolyses the
 CC polypeptide under the conditions; and (ii) the mutant cell produces more
 CC of the polypeptide than the parent cell when both cells are cultivated
 CC under the conditions; and (b) recovering the polypeptide. The method is
 CC useful for producing polypeptides encoded by a DNA sequence which does
 CC not contain the mutation.
 XX
 SQ Sequence 178 AA;
 CC
 CC Query Match 3.5%; Score 7; DB 20; Length 178;
 CC Best Local Similarity 100.0%; Pred. No. 49;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 124 LLLQFIL 140
 CC | | | | | |
 CC Db 104 111qf11 110
 CC
 CC RESULT 14
 CC AAY95242
 CC ID AAY95242 standard; Protein: 245 AA.
 CC XX
 CC AC AAY95242;
 CC XX
 CC DT 26-JUN-2001 (first entry)
 CC XX
 CC DE Human protein sequence SEQ ID NO:17369.
 CC XX
 CC
 CC KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 CC XX Homo sapiens.
 CC PN EP1074617-A2.
 CC XX
 CC PU 07-FEB-2001.
 CC XX
 CC PF 28-JUL-2000; 2000EP-0115156.
 CC XX
 CC PR 25-01-1999; 99JP-0248036.
 CC PR 27-AUG-1999; 99JP-0300234.
 CC PR 11-JAN-2000; 2000JP-0118776.
 CC PR 02-MAY-2000; 2000JP-0183737.
 CC PR 09-JUN-2000; 2000JP-0241849.
 CC XX
 CC (HELI-) HELIX RES INST.
 CC PA
 CC OTa T., Isodai I., Nishikawa I., Hayashi K., Saito K., Yamamoto J;
 CC Ishii S., Sugiyama I., Wakamatsu A., Nagai K., Otsuki I;
 CC WPI: 2001-318749/34.
 CC XX
 CC PR primer sets for synthesizing polynucleotides, particularly the 5602
 CC full-length cDNAs defined in the specification, and for the detection
 CC and/or diagnosis of the abnormality of the proteins encoded by the
 CC full-length cDNAs -
 CC PS
 CC Claim 8: SEQ ID 17369; 2537pp + CD ROM; English.
 CC XX
 CC The present invention describes primer sets for synthesising 5602
 CC full length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides, or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAY93166 to AAY13628 and
 CC AAY13633 to AAY18742 represent human cDNA sequences; AAY92446 to
 CC AAY95893 represent human amino acid sequences; and AAY13629 to AAY13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC XX
 CC SQ Sequence 245 AA;
 CC
 CC Query Match 3.5%; Score 7; DB 22; Length 245;
 CC Best Local Similarity 100.0%; Pred. No. 65;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 16 SENDLEE 22
 CC | | | | | |
 CC Db 41 sendlee 47
 CC
 CC RESULT 15
 CC AAU35868
 CC ID AAU35868 standard; Protein: 254 AA.
 CC XX
 CC AC AAU35868;
 CC XX
 CC DT 14-FEB-2002 (first entry)
 CC XX

DE Helicobacter pylori cellular proliferation protein #181.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Helicobacter pylori.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI: 2001-611495/70.
 DR N-PSDB: AMS53727.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 11461; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: the sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 254 AA:

Query: Match 3.5%, Score 7, 22, Length 254;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 ENDLEEV 23
 Db 15 endleev 21
 |||||

Search completed: July 2, 2002, 16:06:03
 Job time: 180 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 GenComics Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:03:59 / Search time 44.95 seconds
(without alignments)
109,466 Million cell updates/sec

Title: US-09-603-665-5_COPY_1867_2067

Perfect score: 201
Sequence: 1 UAFUFAIDFRAQHSNDL.....CIAQSVAMADDSWKPLNY 201

Scoring table: OLIGO

Gapop 60.0 , gapext 60.0

Searched: 241628 seqs, 24425544 residues

Word size : 6

Total number of hits satisfying chosen parameters: 127

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post processing: Listing first 45 summaries

Database : Issued Patents AA *

1: /seq2_6/prodata/2/1aa/5A_7MM.pep.*
2: /seq2_6/prodata/2/1aa/5A_7MM.pep.*
3: /seq2_6/prodata/2/1aa/5A_7MM.pep.*
4: /seq2_6/prodata/2/1aa/5A_7MM.pep.*
5: /seq2_6/prodata/2/1aa/5A_7MM.pep.*
6: /seq2_6/prodata/2/1aa/5A_7MM.pep.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	3.5	41	US-08-630-915A-79	Sequence 79, Appl
2	7	3.5	98	US-08-211-202-130	Sequence 130, App
3	7	3.5	117	US-08-545-809A-139	Sequence 139, App
4	7	3.5	178	US-08-928-692-57	Sequence 57, Appl
5	7	3.5	428	US-09-347-833-6	Sequence 6, Appl1
6	7	3.5	655	US-09-347-833-4	Sequence 4, Appl1
7	7	3.5	911	US-08-928-692-59	Sequence 59, Appl
8	7	3.5	916	US-08-928-692-58	Sequence 58, Appl
9	7	3.5	1015	US-07-596-467-6	Sequence 6, Appl1
10	7	3.5	1015	US-07-596-467-6	Sequence 6, Appl1
11	7	3.5	1015	US-07-596-467-6	Sequence 6, Appl1
12	6	3.0	34	US-08-178-477B-28	Sequence 28, Appl
13	6	3.0	35	US-08-252-666B-9	Sequence 9, Appl1
14	6	3.0	133	US-09-085-761A-47	Sequence 47, Appl
15	6	3.0	133	US-09-085-761A-47	Sequence 47, Appl
16	6	3.0	131	US-09-087-465-22	Sequence 22, Appl
17	6	3.0	131	US-09-087-465-22	Sequence 22, Appl
18	6	3.0	131	US-09-087-465-22	Sequence 22, Appl
19	6	3.0	178	US-08-477-451-46	Sequence 46, Appl
20	6	3.0	181	US-08-482-142-193	Sequence 193, App
21	6	3.0	181	US-08-482-142-193	Sequence 193, App
22	6	3.0	181	US-08-482-142-193	Sequence 193, App
23	6	3.0	195	US-08-882-704A-2	Sequence 2, Appl1
24	6	3.0	202	US-08-469-486-56	Sequence 56, Appl
25	6	3.0	202	US-08-469-486-56	Sequence 56, Appl
26	6	3.0	202	US-08-469-486-56	Sequence 56, Appl
27	6	3.0	208	US-08-531-525-16	Sequence 16, Appl

Sequence 15, Appl
Sequence 17, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 18, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 6, Appl
Sequence 8, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-630-915A-79
; Sequence 79, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 9309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Diana M.
; APPLICANT: MCGONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.40
; CURRENT APPLICATION DATA: US/08/630,915A
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 1E,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 740-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-630-915A-79

Query Match: 3.5%, Score 7; DB 4; Length 41;
Best Local Similarity: 100.0%, Pctd. No. 3.2;

Matches 7; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 116 ENDPEKC 122

Db 29 ENDPEKC 35

RESULT 2

US-08-211-202-130
; Sequence 130, Application US/08211202
; Patent No. 5565332
; GENERAL INFORMATION:
; APPLICANT: HUGHENBOM, Hendricus Rencius Jacobus Mattheus
; APPLICANT: BAUER, Michael
; APPLICANT: JESPERS, Laurent Stephane Anne Therese
; APPLICANT: WINTER, Gregory Paul
; TITLE OF INVENTION: Production of chimeric antibodies - a
; TITLE OF INVENTION: combinatorial approach
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David W. Clough, Marshall C. Loele Gerstein Murray &
; ADDRESSEE: Boran
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (RPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,202
; FILING DATE: 23-SEP-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GR 9120252 3
; FILING DATE: 23-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9120377.8
; FILING DATE: 25-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GR 9206318.9
; FILING DATE: 24-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GR 9206372.6
; FILING DATE: 24-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GR92/00883
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: David W. Clough
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 28111/31960
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25 3656
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-211-202-130

Query Match 3.5%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KQPRVT 176

Db 63 KQPRVT 69

RESULT 3

US-08-545-809A-139
; Sequence 139, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Fumihiko
; APPLICANT: HANJO, Tasuku
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 139:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-545-809A-139

Query Match 3.5%; Score 7; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KQPRVT 176

Db 82 KQPRVT 88

RESULT 4

US-08-928-b92-57
; Sequence 57, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Iamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727G No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue

CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION NUMBER: 60/092,415
APPLICANT: Microsoft Office 97
FILING DATE: 12-SEPT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4944,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 178 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5958727e
US-08-928-692-57

Query Match 3.5% Score 7; DB 2; Length 178;
Best Local Similarity 100.0% Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 LLLQFLL 140
DB 104 LLLQFLL 110

RESULT 5

US-09-447-833-6
Sequence 6, Application US/09/447833
Patent No. 6294658
GENERAL INFORMATION:
APPLICANT: Famodu, Layo O.
APPLICANT: Odell, Joan T.
TITLE OF INVENTION: Factors Involved in Gene Expression
FILE REFERENCE: HH-1174
CURRENT APPLICATION NUMBER: US/09/447,833
CURRENT FILING DATE: 1999-07-02
EARLIER APPLICATION NUMBER: 60/092,415
EARLIER FILING DATE: July 10, 1998
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Microsoft Office 97
SEQ ID NO 6
LENGTH: 428
TYPE: PRT
ORGANISM: Glycine max
US-09-447-833-6

Query Match 3.5% Score 7; DB 4; Length 428;
Best Local Similarity 100.0% Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 PLVIOLE 162
DB 457 PLVIOLE 363

RESULT 6

US-09-447-833-4
Sequence 4, Application US/09/447833

Patent No. 6294658
GENERAL INFORMATION:
APPLICANT: Famodu, Layo O.
APPLICANT: Odell, Joan T.
TITLE OF INVENTION: Factors Involved in Gene Expression
FILE REFERENCE: HH-1172
CURRENT APPLICATION NUMBER: US/09/447,833
CURRENT FILING DATE: 1999-07-02
EARLIER APPLICATION NUMBER: 60/092,415
EARLIER FILING DATE: July 10, 1998
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 655
TYPE: PRT
ORGANISM: Oryza sativa
US-09-447-833-4

Query Match 3.5% Score 7; DB 4; Length 655;
Best Local Similarity 100.0% Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 PLVIOLE 162
DB 583 PLVIOLE 589

RESULT 7

US-08-928-692-59
Sequence 59, Application US/38928692
Patent No. 5958727
GENERAL INFORMATION:
APPLICANT: Brody, Howard
APPLICANT: Yaver, Deborah S.
APPLICANT: Lamsa, Michael
APPLICANT: Hansen, Kim
TITLE OF INVENTION: Methods for Modifying the Production of
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5958727 No. 5958727disk of No. 5958727th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/08/928,692
FILING DATE: 12-SEPT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4944,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 911 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5958727e
US-08-928-692-59

```

Query Match          3.5%; Score 7; DB 2; Length 911;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 LLLQFIL 130
DB 637 LLLQFIL 643

RESULT 8
US-08-928-692-58
; Sequence 58, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSER: No. 5958727, No. 5958727, disk of No. 5958727, America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944,200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878 9655
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 916 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727*
US-08-928-692-58

Query Match          3.5%; Score 7; DB 2; Length 916;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 LLLQFIL 130
DB 642 LLLQFIL 648

RESULT 9
US-07-596-467-6
; Sequence 6, Application US/07596467
; Patent No. 5306862
; GENERAL INFORMATION:
; APPLICANT: Chappell, J.
; APPLICANT: Saunders, Court A.
; APPLICANT: Cuellar, R.
; APPLICANT: Wolf, Fred R.
; TITLE OF INVENTION: Method and Composition for Increasing

```

```

; TITLE OF INVENTION: Sterol Accumulation in Higher Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Dressler, Goldsmith, Shore, Sutker & Milnamow
; STREET: 180 N. Statson St
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/596,467
; FILING DATE: 19901012
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 616-5400
; TELEFAX: 312 616-5460
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-596-467-6

Query Match          3.5%; Score 7; DB 1; Length 1045;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLVKPEA 97
DB 10 HLVKPEA 16

RESULT 10
US-07-934-374-6
; Sequence 6, Application US/07934374
; Patent No. 5349126
; GENERAL INFORMATION:
; APPLICANT: Chappell, J.
; APPLICANT: Saunders, Court A.
; APPLICANT: Cuellar, R.
; APPLICANT: Wolf, Fred R.
; TITLE OF INVENTION: Process and Composition for Increasing
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Dressler, Goldsmith, Shore, Sutker & Milnamow
; STREET: 180 N. Statson St
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/934,374
; FILING DATE: 19920814
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 312 616 5400
 TELEFAX: 312 616 5460
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1045 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-944 474 6

Query Match 3.58; Score 7; DB 1; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLYKPEA 97
 |||||
 Db 10 HLYKPEA 16

RESULT 11
 US-07-783-861C-6
 : Sequence 6, Application US/07784861C
 : Patent No. 5460949
 : GENERAL INFORMATION:
 : APPLICANT: Saunders, Court A.
 : APPLICANT: Wolf, Fred R.
 : APPLICANT: Mukharji, Indrani
 : TITLE OF INVENTION: A Method and Composition for Increasing
 : TITLE OF INVENTION: the Accumulation of Squalene and Specific Sterols in
 : TITLE OF INVENTION: Yeast
 : NUMBER OF SEQUENCES: 24
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Amoco Corp., Patents and Licensing Dept.
 : STREET: 200 East Randolph St.
 : CITY: Chicago
 : STATE: Illinois
 : COUNTRY: USA
 : ZIP: 60680-0704
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent It Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/07/783,861C
 : FILING DATE: 19911028
 : CLASSIFICATION: 4-5
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/513,380
 : FILING DATE: 15-NOV-1990
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Galloway, No. 5460949vall B.
 : TELEPHONE: 312 856-7180
 : TELEFAX: 312 856-4972
 : INFORMATION FOR SEQ ID NO: 6:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1045 amino acids
 : TYPE: AMINO ACID
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-07-783 861C-6

Query Match 3.58; Score 7; DB 1; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLYKPEA 97
 |||||
 Db 10 HLYKPEA 16

RESULT 12
 US-08-178-477B-28
 : Sequence 28, Application US/08178477B
 : Patent No. 5756343
 : GENERAL INFORMATION:
 : APPLICANT: WU, CARL; CLOS, JOACHIM;
 : APPLICANT: WILWOOD, J. TIMOTHY; RABINURAN, SKI HAR
 : TITLE OF INVENTION: CELL STRESS
 : TITLE OF INVENTION: TRANSCRIPTIONAL FACTORS
 : NUMBER OF SEQUENCES: 42
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: MORGAN & FINNEGAN
 : STREET: 345 PARK AVENUE
 : CITY: NEW YORK
 : STATE: NEW YORK
 : COUNTRY: USA
 : ZIP: 10154
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: FLOPPY DISK
 : COMPUTER: IBM PC COMPATIBLE
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: WORDPERFECT 5.1
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/178,477B
 : FILING DATE: 07-JAN-1994
 : CLASSIFICATION: 530
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US/07/617,910
 : FILING DATE: 26-NOV-1990
 : CLASSIFICATION: 530
 : ATTORNEY/AGENT INFORMATION:
 : NAME: CAROL M. GRUPPI
 : REGISTRATION NUMBER: 37,341
 : REFERENCE/DOCKET NUMBER: 2026-4103US1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 758-4800
 : TELEFAX: (212) 751-6849
 : TELEX: 421792
 : INFORMATION FOR SEQ ID NO: 28:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 34
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: peptide
 : US-08-178-477B-28

Query Match 3.08; Score 6; DB 1; Length 34;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LFEVCK 25
 |||||
 Db 1 LFEVCK 6

RESULT 13
 US-08-252-966B-9
 : Sequence 9, Application US/03252966B
 : Patent No. 5624818
 : GENERAL INFORMATION:
 : APPLICANT: Eisenman, Robert N.
 : APPLICANT: Hurlin, Peter J.
 : APPLICANT: Ayer, Donald C.
 : TITLE OF INVENTION: Regulatory Proteins that Dimerize with
 : TITLE OF INVENTION: Mad or Max
 : NUMBER OF SEQUENCES: 19
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Christensen, O'Connor, Johnson, and Kindness PLLC
 : STREET: 1420 Fifth Ave., Suite 2800
 : CITY: Seattle

STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/25, 965A
FILING DATE: 01-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/INVENTOR NUMBER: EP-617594
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-8100
TELEFAX: (206) 224-0779
INFORMATION FOR SEQ ID NO. 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
DESCRIPTION: MAL-1 peptide; see Figure 15A
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-252-965A-9

Query Match 3.0%; Score 6; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 EKLKGL 84
| | | | |
Db 22 EKLKGL 27

RESULT 14
US-09-085-761A-47
; Sequence 47, Application US/09085761A
; Patent No. 6335178
; GENERAL INFORMATION:
; APPLICANT: Weiner, Joel H.
; APPLICANT: Turner, Raymond J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN
; NUMBER OF INVENTIONS: SECRETION
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,761A
; FILING DATE: 28-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: UA18-03356

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO. 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-085-761A-47

Query Match 3.0%; Score 6; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AKTEDA 62
| | | | |
Db 90 AKTEDA 95

RESULT 15
US-09-087-465-22
; Sequence 22, Application US/09087465A
; Patent No. 6160092
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Uwe
; APPLICANT: Chen, Xiaomin
; APPLICANT: Darnell Jr., James E.
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE CORP. PORTION OF A STAT AND METHODS OF
; FILE REFERENCE: 600-1-229
; CURRENT APPLICATION NUMBER: US/09/087,465A
; CURRENT FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-087-465-22

Query Match 3.0%; Score 6; DB 4; Length 107;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 FISKER 149
| | | | |
Db 5 FISKER 10

Search completed: July 2, 2002, 16:04:00
Job time: 57 sec

GenDate version 4.5
Copyright (c) 1994 - 2000 CompuGen Ltd.

protein search, using SW model

Run on: July 2, 2002, 16:15:49 / Search time 57.09 Seconds
(without alignments)
338,407 Million cell updates/sec

Title: us-09-603-665-5_copy_1867_2067

Perfect score: 201
Sequence: 1 LTAFLEA 201

Scoring table: 0.150
Gapop 60.0 / Gapext 60.0

Searched: 28338 seqs, 96389334 residues

Word size: 6

Total number of hits satisfying chosen parameters: 710

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: PIR 71:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	8	4.0	467	2 AG0546	probable terminal
2	8	4.0	488	2 T47273	cyanide insensitive
3	8	4.0	875	2 T40023	hypothetical prote
4	8	4.0	935	2 T29393	hypothetical prote
5	7	3.5	98	2 S26914	lg heavy chain V r
6	7	3.5	230	2 T12480	hypothetical prote
7	7	4.5	242	2 J07706	beta crystallin B1
8	7	3.5	251	2 H87692	conserved hypotet
9	7	3.5	254	2 E64716	conserved hypotet
10	7	3.5	276	2 AG3305	23S ribosomal RNA
11	7	3.5	296	2 A87411	hypothetical prote
12	7	3.5	358	2 J05964	apoptosis inhibito
13	7	3.5	359	2 T29550	hypothetical prote
14	7	3.5	372	2 AE3184	alcohol dehydrogen
15	7	3.5	376	2 C19470	hypothetical prote
16	7	3.5	397	2 J07675	acetoacetyl-CoA re
17	7	3.5	400	2 C82213	conserved hypotet
18	7	3.5	414	1 H69963	DNA-damage repair
19	7	3.5	422	2 H87550	hypothetical prote
20	7	3.5	443	2 F84759	cytochrome d (hd-1
21	7	3.5	453	2 F86846	ABC transporter pe
22	7	3.5	537	2 A38170	cytochrome d compl
23	7	3.5	584	2 S77647	mobilisation prote
24	7	3.5	601	2 T03633	hypothetical prote
25	7	3.5	639	1 F69280	iron (II) transpor
26	7	3.5	651	2 T06979	polyadenylate-bind
27	7	3.5	847	2 F84517	probable penicillin
28	7	3.5	875	2 T29429	hypothetical prote
29	7	3.5	911	2 J06016	chitin synthase (E

ALIGNMENTS

RESULT 1

AG0546

probable terminal oxidase chain 1 [imported] - Salmonella enterica subsp. enterica se
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi

C:Date: 00-Nov-2001 #sequence_revision 09 Nov 2001 #real_habdate 27 Nov 2001

C:Accession: AG0546

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
Th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, K.M.; Dowd, L.; White, N.; Farr
S.; Mout, S.; O'Gaara, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PM.D:11677608

A:Accession: AG0546

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-467 <PAR>

A:Cross-references: GB:AL513383; PDB:1AD501629; GSPDB:GN00176

C:Genetics:

A:Gene: SIV0392

C:Superfamily: cytochrome d complex terminal oxidase chain 1

Query Match 4.0% Score 8; DB 2; Length 467;

Best Local Similarity 100.0%; Pred. No. 4.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFLEA 8

|||||||

Db 102 LTAFLEA 109

RESULT 2

T47273

cyanide insensitive terminal oxidase chain ctoA [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa

C:Date: 20-Apr-2000 #sequence_revision 20 Apr 2000 #real_habdate 21 Dec 2000

C:Accession: T47273; G83155

R:Cunningham, L.; Pitt, M.; Williams, H.D.

Mol. Microbiol. 24, 579-591, 1997

A:Title: The ctoAB genes from Pseudomonas aeruginosa code for a novel cyanide-insensit

A:Reference number: Z24440; M0.D:9752403

A:Accession: T47273

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-488 <GUN>

A:Cross-references: EMRL:Y10520; NID:12288964; PDB:1A7555.1; PDB:12208964

A:Experimental source: strain PA01, substrain PA06049

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, F.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen

A:Reference number: AK2450; MIM: 2043737

A:Accession: G83155

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-488 <STO>

A:Cross-references: GB AF004810; GR AK004104; NIH: g950106; PIR: AA07417.1; GSM: G80001

A:Experimental source: strain PA01

A:Gene: ctoA; PA3930

A:Superfamily: cytochrome d complex terminal oxidase chain I

Query Match 4.0%; Score 8; DB 2; Length 488;

Best Local Similarity 100.0%; Pred. No. 4 2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFFLEA 8

|||||

Db 103 LTAFFLEA 110

RESULT 3

T30023 hypothetical protein K08F11.2 *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

A:Accession: T30023

R:Murray, J.; Wohlschlag, P.

submitted to the EMBL Data Library, September 1996

A:Description: The sequence of *C. elegans* cosmid K08F11.

A:Reference number: 220723

A:Accession: T30023

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-875 <MUR>

A:Cross-references: EMBL: U70855; PIR: AA09160.1; GSM: G800022; CESP: K08F11.2

A:Experimental source: strain Bristol N2; clone K08F11

C:Genetics:

A:Gene: CESP:K08F11.2

A:Map position: 4

A:Introns: 97/3; 172/3; 837/2; 875/2

Query Match 4.0%; Score 8; DB 2; Length 875;

Best Local Similarity 100.0%; Pred. No. 7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ENDLEEVG 24

|||||

Db 696 ENDLEEVG 703

RESULT 4

T2480

hypothetical protein K08D10.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

A:Accession: T29390

R:Geisel, C.; Bradshaw, H.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of *C. elegans* cosmid K08D10.

A:Reference number: 220616

A:Accession: T29390

A:Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-935 <GEI>

A:Cross-references: EMBL: M55857; PIR: AAA9805.1; GSM: G800022; CESP: K08D10.1

A:Experimental source: strain Bristol N2; clone K08D10

C:Genetics:

A:Gene: CESP:K08D10.1

A:Map position: 4

A:Introns: 40/2; 143/3; 218/3; 883/2; 923/3

Query Match 4.0%; Score 8; DB 2; Length 935;

Best Local Similarity 100.0%; Pred. No. 7.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ENHLEEVG 24

|||||

Db 742 ENDLEEVG 749

RESULT 5

S26914

Ig heavy chain V region (DP-2) - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000

A:Accession: S26914

R:Tomlinson, I.M.; Walter, G.; Marks, J.D.; Lilwelyn, M.H.; Winter, G.

J. Mol. Biol. 237, 776-798, 1992

A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups o

A:Reference number: S26885; MIM: 9302117

A:Accession: S26914

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-98 <TOM>

A:Cross-references: EMBL: Z12304; NID: g30863; PIR: CAA78174.1; PIR: g33864

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 3.5%; Score 7; DB 2; Length 98;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KFOERVV 176

|||||

Db 63 KFOERVV 69

RESULT 6

T12480

hypothetical protein DKFZp564p2062.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 22-Jun-2000

A:Accession: T12480

R:Plum, H.; Hantsch, S.; Meves, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z17526

A:Accession: T12480

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-230 <PLU>

A:Cross-references: EMBL: AL080096

A:Experimental source: fetal brain, clone DKFZp564p2062

C:Genetics:

A:Note: DKFZp564p2062.1

C:Superfamily: succinate-CoA ligase (ATP-forming) beta chain

Query Match 3.5%; Score 7; DB 2; Length 230;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 MVVKLSR 43

|||||

Db 206 MVVKLSR 212

RESULT 7

JC7706

beta crystallin B1 protein - zebra fish

C:Species: Brachydanio rerio (zebra fish)
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: J67706
 R:Chen, J.Y.; Chand, B.E.; Chen, Y.H.; Lin, C.J.F.; Wu, J.L.; Kuo, C.M.
 Biochem. Biophys. Res. Commun. 285, 105-110, 2001
 A:Title: Molecular cloning, developmental expression, and hormonal regulation of zebrafish
 A:Reference number: J67706; MUID:21131298; PMID:11437379
 A:Accession: J67706
 A:Molecule type: mRNA
 A:Residues: 1-242 <HE>
 A:Cross-References: GR:AJ147957

C:Comment: this protein is involved in regulation by growth factors
 C:Superfamily: beta crystallin
 F:20-25/Region: active proline- and alanine-rich motif #status predicted

Query Match 3.5% Score 7; DB 2; Length 252;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 134 YKIFLPD 140
 DB 41 YKIFLPD 47

RESULT 8
 H87692
 conserved hypothetical protein GC3578 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: H87692
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 u, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: H87692; MUID:21173698; PMID:11259647
 A:Accession: H87692
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-251 <STO>
 A:Cross-References: GR:AF005673; NID:g13425320; PIDN:AAK25540.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: GC3578

Query Match 3.5% Score 7; DB 2; Length 251;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 EALDFRA 14
 DB 201 EALDFRA 207

RESULT 9
 E64716
 conserved hypothetical protein HP1573 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 29-Sep-1999
 C:Accession: E64716
 R:Laub, J.F.; White, O.; Koriavade, A.K.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujita, C.; Bowman, C.; Watney, L.
 Nature 388, 549-557, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpik, P.D.; Smith, H.C.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467
 A:Accession: E64716
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-254 <TOW>
 A:Cross-References: GR:AF000656; GR:AF000811; NID:g2314757; PIDN:AAH08633.1; PID:g231475

C:Superfamily: hypothetical protein H10454

Query Match 3.5% Score 7; DB 2; Length 254;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ENDLEEV 23
 DB 15 ENDLEEV 21

RESULT 10
 AG3305
 23S ribosomal RNA methyltransferase (EC 2.1.1.1) [imported] - Brucella melitensis (st
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AG3305
 R:DelVecchio, V.G.; Kapral, V.; Redkar, P.J.; Patra, G.; Mujer, G.; IGS, L.; Ivanov
 ; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, B.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A:Reference number: AD4252; PMID:11756688
 A:Accession: AG3305
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-276 <KUP>
 A:Cross-References: GR:AF008917; PIDN:AAH51610.1; PID:g17982336; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: HME10429
 A:Map position: 1
 C:Keywords: methyltransferase

Query Match 3.5% Score 7; DB 2; Length 276;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 LENRIGS 167
 DB 67 LENRIGS 73

RESULT 11
 A87311
 hypothetical protein GC0498 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: A87311
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
 u, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: A87311
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-290 <STO>
 A:Cross-References: GR:AF005673; NID:g13421677; PIDN:AAK22485.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: GC0498

Query Match 3.5% Score 7; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LFTLFAG 90
 DB 240 LFTLFAG 246

```

RESULT 12
JC5964
apoptosis inhibitor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 02-Sep-2000
C:Accession: JC5964
K:Stehlik, C.; de Martin, R.; Binder, H.R.; Hipp, J.
Biochem Biophys Res Commun 243, 827-832, 1998
A:Title: Cytokine induced expression of porcine inhibitor of apoptosis protein (iap) fam
A:Reference number: JC5964; MUID:98162622
A:Accession: JC5964
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-358 <STE>
A:Cross-references: GB:U79142; NID:q2957174; PIDN:AAC39171.1; FID:q2957175
C:Superfamily: RING finger homology
F:307-351/domain WING finger homology <PRN>

Query Match 3.5%; Score 7; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 KDRLLTF 70
Db 5 KDRLLTF 11

RESULT 13
T22950
hypothetical protein F58G6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22950
P:Lloyd, C.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19641
A:Accession: T22950
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-359 <WTF>
A:Cross-references: EMBL:Z68217; PIDN:CAA92469.1; GSPDB:GN000022; CESP:F58G6.2
A:Experimental source: clone F58G6
C:Genetics:
A:Gene: CESP:F58G6.2
A:Map position: 4
A:introns: 49/3; 114/3; 166/3; 191/1; 248/1; 303/2; 333/3

Query Match 3.5%; Score 7; DB 2; Length 359;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 LLLQFIL 130
Db 111 LLLQFIL 117

RESULT 14
AF3184
alcohol dehydrogenase [benzyl] xyli [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11 Jan 2002 #text_change 01-Feb-2002
C:Accession: AF3184
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Maiks, D.; Chen, L.; Wood, G.F.; Chen, Y.; Woo, T.
Crage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyaivin, T.; Levy, R.; Li, M.; McClell
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AH2577; PMID:11743193

```

```

A:Accession: AF3184
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <KUR>
A:Cross-references: GB:AE009687; PIDN:AAL45891.1; FID:ql7743636; CESP:R-NN0188
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: xyliB
A:Genome: plasmid
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 3.5%; Score 7; DB 2; Length 372;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 DRLLTFY 71
Db 337 DRLLTFY 343

RESULT 15
SI9740
hypothetical protein B - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 12-Feb-1999
C:Accession: SI9740
R:Kelly, D.J.
submitted to the EMBL Data Library, February 1992
A:Reference number: SI9739
A:Accession: SI9740
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <KEL>
A:Cross-references: EMBL:X63974
C:Superfamily: vhl-type dual specificity phosphoprotein phosphatase homology
F:63-182/domain vhl-type dual specificity phosphoprotein phosphatase homology <VHL>

```

```

Query Match 3.5%; Score 7; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ALDFRAQ 14
Db 319 ALDFRAQ 325

```

Search completed: July 2, 2002, 16:15:40
Job time: 757 sec

GeneTools version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein : protein search, using sw model

Run on: July 2, 2002, 16:19:06 : Search time 40.19 seconds
(without alignments)

257,788 Million cell updates/sec

Index: US-09-603-665-5_copy_1867_2067

Perfect score: 201

Sequence: 1 LIAFFRALDPAHQSENDL.....CIAQFSVAMADDSLWKPLNY 201

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 3871950 residues

Word size : 6

Total number of hits satisfying chosen parameters: 255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post processing: List top first 45 summaries

Database : SWISSPROT_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	47.8	2144	1 BP28_HUMAN	Q9H583 homo sapien
2	75	37.3	958	1 HP28_MACEA	Q9H544 macaca fasc
3	7	4.5	358	1 P1A2_PIG	O62640 sus scrofa
4	7	3.5	414	1 YQJH_HACSU	P54545 bacillus su
5	7	3.5	537	1 CYDA_AZOVI	Q09049 azotobacter
6	7	3.5	845	1 CHS3_EXODE	P40602 exophiala d
7	7	3.5	911	1 CHS3_ASPEU	P54267 aspergillus
8	7	3.5	916	1 CHS3_EMENI	Q00757 emericella
9	7	3.5	1015	1 HMD2_YEAST	P12684 saccharomyc
10	7	3.5	1612	1 AUF4_YEAST	Q12675 saccharomyc
11	6	3.0	40	1 PSAT_YAPA	P48117 cyanophora
12	6	3.0	58	1 YEO2_HAEIN	P44177 haemophilus
13	6	3.0	44	1 EX7S_HAEIN	P43914 haemophilus
14	6	3.0	47	1 GLRX_HAFIN	P45242 haemophilus
15	6	3.0	49	1 IATA_ECOLI	O65948 escherichia
16	6	3.0	134	1 CY22_MOUSE	P35174 mus musculu
17	6	3.0	113	1 RL24_MYCICU	P33193 micrococcos
18	6	3.0	125	1 MERR_STRLI	P30446 streptomyces
19	6	3.0	144	1 YQBL_BACSU	P45928 bacillus su
20	6	3.0	145	1 E334_ADE97	P15135 human adeno
21	6	3.0	148	1 SODM_MYCPH	P53648 mycobacteri
22	6	3.0	158	1 V105_FOWPV	Q9J505 fowlpox vir
23	6	3.0	158	1 FRAA_SCHPO	O74831 schizosacch
24	6	3.0	155	1 DEST_CHICK	P18459 gallus gall
25	6	3.0	155	1 DEST_HUMAN	P18282 homo sapien
26	6	3.0	155	1 DEST_MOUSE	Q9905 mus musculu
27	6	3.0	176	1 Y157_ARCFU	P18282 homo sapien
28	6	3.0	148	1 VF59_MPTA	O58954 methanococ
29	6	3.0	146	1 Q1DR_ECOLI	O59431 escherichia
30	6	3.0	194	1 RAB2_DICDI	P36409 dictyostell
31	6	3.0	202	1 TETN_MYCPH	P05452 homo sapien
32	6	3.0	206	1 SODM_MYCPH	Q95419 mycobacteri
33	6	3.0	208	1 VPT_NPVOP	O10418 orgyia pseu

RESULT 1

ID	BP28_HUMAN	STANDARD:	PR1:	2144 AA.
AC	Q9H583: Q9NW23:			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Protein HAP28.			
GN	HAP28.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANTS S-1694; A-1854; D-1967 AND G-2017.			
RA	Bouqueleret L., Chumakov J., Barry G., Cohen-Akeline A.			
RT	"A novel BAP28 gene and protein."			
RL	Patent number WO0100669, 04-JAN-2001.			
RN	[2]			
RP	SEQUENCE OF 1534-2144 FROM N.A.			
RA	Cobley V.			
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1777-2144 FROM N.A.			
RA	Isotai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sudawara M.,			
RA	Wagatsuma M., Asouiri T., Kaku Y., Kodaira H., Kondo H., Sudawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Watanabe S., Kimura K., Mirakami K., Ishii S., Kawai Y., Saito K.,			
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y.,			
RA	Ninomiya K., Iwayanagi T.			
RT	"NDO human cDNA sequencing project."			
RL	Submitted (FEB 2000) to the EMBL/GenBank/DBJ databases.			
OC	-1- SIMILARITY: BELONGS TO THE BAP28 FAMILY.			
OC	-1- SIMILARITY: CONTAINS 1 HEAT REPEAT.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as the content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcements/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL: AX067150; CAC26776.1; ..			
DR	EMBL: AL136105; CAC15948.1; ..			
DR	EMBL: AK001221; BAA01564.1; ALT_INIT.			
DR	InterPro: IPR000457; HEAT_repeat.			
KW	PROSITE; IPR000457; HEAT_repeat.			
KW	Polymorphism.			
FT	REPEAT 2106 2142 HEAT.			
FT	REPEAT 1694 1694 N->S.			
FT	VARIANT 1854 1854 /FTID-VAR_010939.			
FT	VARIANT 1854 1854 V->A.			
FT	VARIANT 1967 1967 /FTID-VAR_010940.			
FT	VARIANT 1967 1967 N->D.			

P49104 703 ways (m
P49104 703 ways (m
P08886 homo sapien
Q05975 lymphoid tis
Q05994 mus musculu
Q05971 erythrocyt
Q05972 rattus norv
Q05970 chlamydomon
P46863 volvox cart
G4741 homo sapien
P48444 serralia ma
P49973 haemophilus

ALIGNMENTS

```

FT VARIANT 2017 2017 /FTID-VAP_010941
FT 2017 2017 E-5-G
SQ SEQUENCE 2144 AA; 242355 MW; D66816E7B0BC9H7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 75; DB 1; Length 958;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPAFFLEALDPAQHSNDLEKVGKTEKNCIDCLVAMVVKUSFVTFPPILFFKLFHWAKTE 60
DB 1867 TPAFFLEALDPAQHSNDLEKVGKTEKNCIDCLVAMVVKUSFVTFPPILFFKLFHWAKTE 1926

QY 61 DAPKDRLLTFYNLADCIAPKIKGCLFTLPAGHLVKKPFDLTL 100
DB 1927 DAPKDRLLTFYNLADCIAPKIKGCLFTLPAGHLVKKPFDLTL 1466

RESULT 2
RP28 MACFA STANDARD; PRT; 958 AA.
AC O9GM44;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Protein HAP28 (Fragment).
GN HAP28.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Primates, Catarrhini, Cercopithecoidea,
OC Cercopithecinae, Macaca.
OX NCBI_TaxID=9541;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Osada N., Hida M., Kusuda J., Tanuma P., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sudano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (Oct-2000) to the FMBI/Genbank/EMBL databases
CC 1- SIMILARITY: BELONGS TO THE HAP28 FAMILY
CC 1- SIMILARITY: CONTAINS 1 HEAT REPEAT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC FMBI: AH049842; HAI16728.1; ALT_INIT.
DR InterPro: IPR000357; HEAT_REPEAT.
DR PROSITE: PS00777; HEAT_REPEAT; FALSE_NEG;
FT NON_TER 1 1
FT REPEAT 920 956 HEAT.
SQ SEQUENCE 958 AA; 108644 MW; 3D0D95C3623CFB31 CRC64;

Query Match
Best Local Similarity 37.3%; Score 75; DB 1; Length 958;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 TNCNIDCLVAMVVKUSFVTFPPILFFKLFHWAKTEAPKDRLLTFYNLADCIAPKIKGCLF 85
DB 706 TNCNIDCLVAMVVKUSFVTFPPILFFKLFHWAKTEAPKDRLLTFYNLADCIAPKIKGCLF 765

QY 86 TLPAGHLVKKPFDLTL 100
DB 766 TLPAGHLVKKPFDLTL 780

```

DT 16-OCT 2001 (Rel. 40, Last annotation update)
DE Hypothetical 47.0 kDa protein in qing-ansu intergenic region.
GN Y03H.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OT Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / JH642;

RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,

RA Sato T., Takeuchi M.;

RA Submitted (May 1996) to the EMBL/GenBank/DBJ databases

CC -1- SIMILARITY: BELONGS TO THE IMPB/MUCB/SAMB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D84442; DAA12614.1;

DR EMBL: Z99116; CAB14319.1;

DR Subtilist: BG11737; Y1JB.

DR InterPro: IPR001126; IMUC.

DR Pfam: PF00817; IMS; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 414 AA: 47020 MW: 982462080D8448F CRC64;

Query Match 3.5%; Score 7; DB 1; Length 414;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 AEKIKGL 84

|||||

Db 193 AEKIKGL 199

RESULT 5

ID CYDA_AZOV1 STANDARD; PRT: 537 AA.

AC Q09049;

DT 01-NOV-1997 (Rel. 35, Created)

DI 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Cytochrome D ubiquinol oxidase subunit I (EC 1.10.3.-)

GN CYDA

OS Azotobacter vinelandii;

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Azotobacter.

OX NCBI_TaxID=354;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CA;

RE MEDLINE=9201187; PubMed=1655703;

RA Moshiri F., Chawla A., Maier R.J.;

RA "Cloning, characterization, and expression in Escherichia coli of the

RT genes encoding the cytochrome d oxidase complex from Azotobacter

RT vinelandii.";

RL J. Bacteriol. 173:6230-6241(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CA;

RE MEDLINE=92078187; PubMed=1660468;

RA Moshiri F., Smith F.G., Tacchino J.P., Maier R.J.;

RA "Transcriptional regulation of cytochrome d in nitrogen fixing

RT Azotobacter vinelandii: Evidence that up regulation during N2

RT fixation is independent of nifA but dependent on ntrA.";

RL J. Biol. Chem. 266:24169-24174(1991).

RN [3]

PP CHARACTERIZATION.
PX MEDLINE=9109879; PubMed=2170336;
RA Kelly M.J.S., Poole R.K., Yates M.G., Kennedy C.;
RT "Cloning and mutagenesis of genes encoding the cytochrome bd terminal
RT oxidase complex in Azotobacter vinelandii: mutants deficient in the
RT cytochrome d complex are unable to fix nitrogen in air.";
RL J. Bacteriol. 172:6010-6019(1990).
CC -1- FUNCTION: MAY BE INVOLVED IN MAINTAINING THE LOW INTRACELLULAR
CC OXYGEN CONCENTRATION REQUIRED FOR NITROGEN FIXATION.
CC -1- CATALYTIC ACTIVITY: Ubiquinol-8 + O(2) = Ubiquinone-8 + H(2)O.
CC -1- COFACTOR: CONTAINS FET PROTOHEME IX CENTER H558.
CC -1- SUBUNIT: HETEROLOGOUS OF SUBUNITS I AND II.
CC -1- SIMILARITY LOCATION: Integral membrane protein. Inner membrane.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M77787; AAA22123.1;

DR EMBL: S57866; AAD14886.1;

DR InterPro: IPR002585; Bac_Ubq_Cox.

DR Pfam: PF01654; Bac_Ubq_Cox; 1.

KW Oxidoreductase, Electron transport; Transmembrane; Inner membrane;

KW Heme.

FT DOMAIN 1 24 CYTOPLASMIC (POTENTIAL);

FT TRANSMEM 25 44 POTENTIAL;

FT DOMAIN 45 96 PERIPLASMIC (POTENTIAL);

FT TRANSMEM 97 116 POTENTIAL;

FT DOMAIN 117 131 CYTOPLASMIC (POTENTIAL);

FT TRANSMEM 132 151 POTENTIAL;

FT DOMAIN 152 189 PERIPLASMIC (POTENTIAL);

FT TRANSMEM 190 209 POTENTIAL;

FT DOMAIN 210 221 CYTOPLASMIC (POTENTIAL);

FT TRANSMEM 222 241 POTENTIAL;

FT DOMAIN 242 394 PERIPLASMIC (POTENTIAL);

FT TRANSMEM 395 414 POTENTIAL;

FT DOMAIN 415 472 CYTOPLASMIC (POTENTIAL);

FT TRANSMEM 473 492 POTENTIAL;

FT DOMAIN 493 537 PERIPLASMIC (POTENTIAL);

FT METAL 188 188 IRON (HEME B558 AXIAL LIGAND) (BY

FT SIMILARITY).

SQ SEQUENCE 537 AA: 59719 MW: 8F78A49F6A49028 CRC64;

Query Match 3.5%; Score 7; DB 1; Length 537;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFPLE 7

|||||

Db 103 LTAFPLE 109

RESULT 6

CHS3_EXCODE

ID CHS3_EXCODE STANDARD; PRT: 885 AA.

AC P30602; 074678;

DT 01-APR-1993 (Rel. 25, Created)

DI 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Chitin synthase 3 (EC 2.4.1.15) (Chitin-3BP acetyl-glucosaminyl

DE transferase 3) (Class-III chitin synthase 3).

GN CHS3.

OS Exophiala dermatitidis (Nanqiella dermatitidis).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothyriales;

OC Chaetothyriales; Herpotrichiellaceae; mitosporic Herpotrichiellaceae;

OC Exophiala.

OX NCBI_TaxID=5970;

RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-8656;
RA Wang Z., Graybill J.P., Szaniszlo P.J.;
RT "Characterization and expression studies of WUCHS3, a gene that
RT encodes a class III chitin synthase and contributes to virulence in
RT Wangiella (Exophiala) dermatitidis.";
RL Submitted (MAP-1998) to the EMBL/GenBank/DBJ databases
[2]
RP SEQUENCE OF 214-410 FROM N.A.
RX MEDLIN-92115692; PubMed-1731323;
RA Brown A.P., Chen-Wu L.L., Mennery M., Young R., Szaniszlo P.J.,
RA Robbins P.W.;
RT "Classification of fungal chitin synthases.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992)
CC -!- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + ((1,4)-(N-acetyl-
CC beta-D-glucosaminyl))-(N) = UDP + ((1,4)-(N-acetyl-beta-D-
CC glucosaminyl))-(N+1);
CC -!- SUBCELLULAR LOCATION: Plasma membrane-bound.
CC -!- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
CC III.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X94244; CAA63928.1; -
CC EMBL: U39478; AAB07679.1; -
CC EMBL: U39479; AAB07679.1; -
CC Interpro: IPR002923; Chitin_synth.
CC Pfam: PF01644; Chitin_synth; 1.
CC Prodom: P0002998; Chitin_synth; 1.
CC Prodom: P0002998; Chitin_synth; 1.
CC Transferrase; Glycosyltransferase; Transmembrane; Cell wall;
CC Multigene family.
CC TRANSMEM 579 599 POTENTIAL.
CC TRANSMEM 624 644 POTENTIAL.
CC TRANSMEM 659 679 POTENTIAL.
CC TRANSMEM 711 731 POTENTIAL.
CC TRANSMEM 840 860 POTENTIAL.
CC TRANSMEM 879 899 POTENTIAL.
CC CONFLICT 334 337 PPVQ -> LPSK (IN REF. 1).
CC CONFLICT 537 537 R -> C (IN REF. 2; AAB07679).
CC CONFLICT 628 629 IV -> MM (IN REF. 2; AAB07679).
CC CONFLICT 892 892 F -> N (IN REF. 2; AAB07679).
SQ SEQUENCE 911 AA; 101669 MW; BC893F8A00BE48D CRC64;

Query Match 3.5% Score 7; DB 1; Length 911;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 124 LLLQFIL 130
Db 633 LLLQFIL 639

RESULT 7
CHSG_ASPEU
ID CHSG_ASPEU STANDARD; PRT: 911 AA.
AC P54267; Q09031; Q09032.
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chitin synthase G (EC 2.4.1.16) (Chitin UDP acetyl-glucosaminyl
DE transferase G) (Class-III chitin synthase G).
GN CHSG.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Mitosporie Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
[1]
RP SEQUENCE FROM N.A.
RX MEDLIN-96347138; PubMed-8736545;
RA Meliado E., Aufavre-Brown A., Gow N.A.P., Holden D.W.;
RT "The Aspergillus fumigatus chsg and chsg genes encode class III
RT chitin synthases with different functions.";
RL Mol. Microbiol. 20:667-679(1996)

```

RL Biosci. Biotechnol. Biochem. 58:1828-1835(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCSC 89;
 RA Aspergillus nidulans chitin synthase B gene.*;
 RT Submitted (FEB 1997) to the EMBL/GenBank/DBJ databases.
 CC 1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOSYNTHESIS
 CC 1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + (1,4)-(N-acetyl-
 CC beta-D-glucosaminyl)(N) - UDP + (1,4)-(N-acetyl-beta-D-
 CC glucosaminyl)(N).
 CC 1- SUBCELLULAR LOCATION: plasma membrane-bound.
 CC 1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
 CC III.
 CC -----
 CC this swiss prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D21284; BA04807.1;
 DR EMBL: D83216; BA01845.1;
 DR InterPro: IPR002923; Chitin synth.
 DR Pfam: PF01644; Chitin synth. 1.
 DR ProDom: PD02498; Chitin synth. 1.
 DR transferase; glycosyltransferase; transmembrane; Cell wall;
 KW Multigene family.
 KW SEQUENCE 918 AA; 102566 MW; 08FAA9647FF6A1F7 CPr54;
 SQ
 Query Match 3.5%; Score 7; DB 1; Length 916;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 124 LILOFIL 190
 DB 642 LILOFIL 648
 RESULT 9
 HMD2_YEAST STANDARD; PRT; 1045 AA.
 A' P12684;
 DT 01-OCT-1989 (Ref. 12, Created)
 DI 01-OCT-1989 (Ref. 12, Last sequence update)
 DI 01-NOV-1997 (Ref. 35, Last annotation update)
 DE 4-hydroxy-3-methylglutaryl-coenzyme A reductase 2 (EC 1.1.1.14) (HMG-
 DE CoA reductase 2).
 GN HMG2 OR YLR450W OR Y1924.2.
 GS Saccharomyces cerevisiae (baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID:4912;
 RN 111
 RP SEQUENCE FROM N.A.
 RX MEDLINE:6912721; PubMed:1055635;
 RA Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J.;
 RT structural and functional conservation between yeast
 RT hydroxy-3-methylglutaryl coenzyme A reductases, the ra-
 RT tional enzyme of sterol biosynthesis.*;
 RL Mol. Cell. Biol. 8:3797-3808(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288c / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Dittmer D., Du Z.,
 RA Jovello A., Fulton L., Gattung S., Green E., Kirsten J.,
 RA Kueba T., Hallsworth K., Hawkins J., Hillier J., Jier M.,
 RA Johnson D., Johnston J., Langston Y., Latreille P., Le P.,
 RA Mardis E., Mezezes S., Miller N., Nhan M., Parley A., Poulos D.,
 RA Rikken L., Riles L., Taich A., Trevasxis S., Vignati D.,
 RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
 RL Submitted (MAR 1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 772-961 FROM N.A.
 RX MEDLINE:86287298; PubMed:1526336;
 RA Basson M.E., Thorsness M., Rine J.;
 RT "Saccharomyces cerevisiae contains two functional genes encoding 4-
 RT hydroxy-3-methylglutaryl-coenzyme A reductase.*";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5563-5567(1986).
 CC 1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
 CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
 CC ENZYME OF THE STEROL BIOSYNTHESIS.
 CC 1- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) - (S)-4-
 CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.
 CC 1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
 CC 1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC 1- SIMILARITY: BELONGS TO THE HMG-CoA REDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M22255; AAA34677.1;
 DR EMBL: U23382; AAG67527.1;
 DR PIR: B24317; B24317.
 DR PIR: B30249; R0239.
 DR SGD: S0004442; HMG2.
 DR InterPro: IPR002202; HMG-CoA red.
 DR InterPro: IPR000731; HMGCRatched_5PM.
 DR Pfam: PF00368; HMG-CoA red. 1.
 DR PRINTS: PR00071; HMGCOARDTASE.
 DR PROSITE: PS00066; HMG-CoA_REDUCTASE_1; 1.
 DR PROSITE: PS00318; HMG-CoA_REDUCTASE_2; 1.
 DR PROSITE: PS01192; HMG-CoA_REDUCTASE_3; 1.
 DR PROSITE: PS50065; HMG-CoA_REDUCTASE_4; 1.
 DR PROSITE: PS50156; SSD; 1.
 DR Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
 KW Cholesterol biosynthesis; NADP; Multigene family.
 FT DOMAIN 1 523
 FT DOMAIN 523 1045
 FT LINKER.
 FT CATALYTIC.
 FT TRANSMEM 27 53
 FT TRANSMEM 186 210
 FT TRANSMEM 241 265
 FT TRANSMEM 299 323
 FT TRANSMEM 331 356
 FT TRANSMEM 397 421
 FT TRANSMEM 498 523
 FT ACT_SITE 710 710
 FT ACT_SITE 920 920
 FT ACT_SITE 1016 1016
 FT ACT_SITE 115 115
 FT CARBOHYD 150 150
 FT CARBOHYD 158 158
 FT CARBOHYD 179 179
 FT CARBOHYD 428 428
 FT CARBOHYD 455 455
 FT CARBOHYD 613 613
 FT CARBOHYD 774 774
 FT CARBOHYD 885 885
 FT CARBOHYD 1045 1045
 FT SEQUENCE 1045 AA; 115691 MW; 1FD9DCD4AC01B15E CPr54;
 SQ
 Query Match 3.5%; Score 7; DB 1; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 91 HLKPKPA 97

```

DB 10 HAVKPA 16
|||||
RESULT 10
ATCA_YEAST
ID ATCA_YEAST STANDARD; PRT: 1612 AA.
AC Q12675;
DT 01-NOV-1997 (rel 35, created)
DT 01-NOV-1997 (rel 35, last sequence update)
DT 01-MAR-2002 (rel 41, last annotation update)
DE YP0093W CP Y08557.01.
GN YP0093W CP Y08557.01.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 111
RP
RC STRAIN=S288C / AB972;
RA Murphy L., Harris D., Barrell R.G., Rajandream M.A.;
PL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases
CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
CC OF ATP COUPLED WITH THE TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (EI-E2 ATPASES) SUBFAMILY IV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: Z47746; CAAB7668.1;
CC SGD: S0002500; YDR093W.
CC InterPro: IPR001757; EI-E2-ATPase.
CC InterPro: IPR001454; Hydrolase.
CC Pfam: PF00702; Hydrolase, 1.
CC PROSITE: PS00154; ATPASE_EI_E2; 1.
CC Hypothetical protein, Hydrolase, Transmembrane; Phosphorylation;
KW Magnesium; ATP-binding.
FT MCD_RES 712 712 PHOSPHORYLATION (PROBABLE).
SQ SEQUENCE 1612 AA: 184617 MW: 997106.56455889 GPGK4.

Query Match 3.5%; Score 7; DB 1; Length 1612;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LIPYNIA 74
|||||
DB 1265 LIPYNIA 1271

RESULT 11
PSAJ_CYPAA
ID PSAJ_CYPAA STANDARD; PRT: 40 AA.
AC P48117;
DT 01-FEB-1996 (rel 33, created)
DT 01-FEB-1996 (rel 33, last sequence update)
DT 01-FEB-1996 (rel 41, last annotation update)
DE Photosystem I reaction center subunit IX (PSI-I).
GN PSAJ.
OS Cyanophora paradoxa.
OC Cyanophyta; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OC Eukaryota.
OX NCBI_TaxID=2762;
RN 111
RP
RC STRAIN=LB555 / PRINGSHEIM;
RA Bryant D.A.;
PL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases
CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
CC OF ATP COUPLED WITH THE TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (EI-E2 ATPASES) SUBFAMILY IV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: U0821; AA01183.1;
CC Mendeel, 7884; CYAPA.PSAJ.1.
CC InterPro: IPR002615; PSI_PSAJ.
CC Pfam: PF01701; PSI_PSAJ; 1.
CC ProDom: PD004198; PSI_PSAJ; 1.
CC Cyanelle, Photosystem I. Photosynthesis; Transmembrane.
KW TRANSMBM 6 26 POTENTIAL.
FT TRANSMEM 6 26
SQ SEQUENCE 40 AA: 4481 MW: 79848CA3bAU07153 CKC64;

Query Match 3.0%; Score 6; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIAFFL 6
|||||
DB 16 LIAFFL 21

RESULT 12
YF02_HAFIN
ID YF02_HAFIN STANDARD; PRT: 68 AA.
AC P44177;
DT 01-NOV-1995 (rel 32, created)
DT 01-NOV-1995 (rel 32, last sequence update)
DT 16-OCT-2001 (rel 40, last annotation update)
DE Hypothetical protein H11402.
GN H11402.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN 111
RP
RC STRAIN=RD / KW20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty R.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., H-Down F., Cotton M.D.,
RA Uetrichback T.R., Hanna M.C., Nguyen D.T., Sandek D.M., Brandon P.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Googhegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

```


RA Venter J.C.
 RT "Whole genome random sequencing and assembly of Haemophilus
 RL influenzae Rd."
 RL Science 269:496-512(1995).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: U32829; AAC23182.1;
 TIGR: H11437;
 DR InterPro: IPR002109; Exonuc_VII_S.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 68 AA: 7841 MW: 396344AA8UB7D29 CRC64:

Query Match 3.0%, Score 6, DH 1, Length 68;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 DRLLIF 70
 DB 43 DRLLIF 48

RESULT 14
 EX7S_HAEIN STANDARD: PRT: 84 AA.
 AC P45144;
 DT 01-NOV-1995 (Rel. 42, Created)
 DT 01-NOV-1995 (Rel. 42, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exodeoxyribonuclease VII small subunit (EC 3.1.11.6) (Exonuclease VII
 DE small subunit).
 DE XSXB OR H11437.
 GN Haemophilus influenzae.
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID:727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN RD / KW20 / ATCC 51907.
 RX MEDLINE-95350630; PubMed-7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-L., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Hiteckack T.F., Hanna M.C., Nguyen G.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geohagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.
 RT "Whole genome random sequencing and assembly of Haemophilus
 RL influenzae Rd."
 RL Science 269:496-512(1995).

CC -1- FUNCTION: BIDIRECTIONALLY DEGRADATES SINGLE-STRANDED DNA INTO 1-3' 5'
 CC ACID INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
 CC INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES. IT CAN ALSO DEGRADE 3'
 CC OR 5' SS REGIONS EXTENDING FROM THE TERMINI OF DUPLEX DNA
 CC MOLECULES AND DISPLAYED SS REGIONS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage is either 5' to 3'
 CC or 3' to 5'-direction to yield 5'-phosphonucleotides
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF LARGE AND SMALL SUBUNITS (BY
 CC SIMILARITY)
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE XSXB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: U32829; AAC23182.1;
 TIGR: H11437;
 DR InterPro: IPR002109; Exonuc_VII_S.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 84 AA: 9511 MW: 4A3C6B1F055BA74E CRC64:

Query Match 3.0%, Score 6, DH 1, Length 84;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 KTEDAP 63
 DB 66 KTEDAP 71

RESULT 14
 GLRX_HAEIN STANDARD: PRT: 87 AA.
 AC P45242;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutaredoxin.
 DE GRXA OR GRX OR H11532.
 GN Haemophilus influenzae.
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID:727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN RD / KW20 / ATCC 51907;
 RX MEDLINE-95350630; PubMed-7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-L., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Hiteckack T.F., Hanna M.C., Nguyen G.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geohagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.
 RT "Whole genome random sequencing and assembly of Haemophilus
 RL influenzae Rd."
 RL Science 269:496-512(1995).

CC -1- FUNCTION: HAS A GLUTATHIONE-DISULFIDE OXIDOREDUCTASE ACTIVITY IN
 CC THE PRESENCE OF NADPH AND GLUTATHIONE REDUCTASE. REPORTS LOW
 CC MOLECULAR WEIGHT DISULFIDES AND PROTEINS (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE GLUTAREDOXIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: U32829; AAC23182.1;
 TIGR: H11532;
 DR InterPro: IPR002109; Glutaredoxin.
 DR InterPro: IPR000063; Thioredoxin.
 DR Pfam: PF00462; glutaredox n: 1.
 DR PRINTS: PR00160; GLUTAREDOXIN.

PA Apodara I., Anantharaman T.S., Lin D., Yen G., Schwartz D.C.,
PA Weich P.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
PL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / PMID 10509952;
RX MEDLINE=21156233; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida F., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
PL DNA Res. 8:11-22(2001).
RN [7]
RP REVIEW.
RX MEDLINE=20117987; PubMed=10652088;
RA Herks R.C., Sargent F., Palmer T.;
RT "The Tat protein export pathway";
PL Mol. Microbiol. 35:260-274(2000).
CC -!- FUNCTION: EQUIPPED FOR APPROPRIATE LOCALIZATION OF PRECURSOR PROTEINS
CC REAPING SIGNAL PEPTIDES WITH THE TWIN ARGININE CONSERVED MOTIF
CC S/T-P-R-X-F-L-K. THIS SPEC-INDEPENDENT PATHWAY IS TERMED TAT FOR
CC 'TWIN-ARGININE' TRANSLLOCATION SYSTEM. THIS SYSTEM MAINLY TRANSPORTS
CC PROTEINS WITH BOUND COFACTORS THAT REQUIRE FOLDING PRIOR TO
CC EXPORT.
CC -!- SUBCELLULAR LOCATION: INNER-MEMBRANE BOUND (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE TATAZE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR PMRL: AT005830; CAA06724 1; -
DR EMBL: AF067848; AAC19240 1; ALT_INIT.
DR EMBL: M87049; AAN57533 1; ALT_SEQ.
DR EMBL: AB000459; AAC76839 1; ALT_INIT.
DR EMBL: AB000614; AAG54042 1; ALT_INIT.
DR EMBL: AF002567; BAB38189 1; -
DR PIR: S30727; S30727.
DR EcoGene: EG11478; tata.
DR InterPro: IPR003369; Meta-Hcfl06.
DR Pfam: PF02416; Meta-Hcfl06; 1.
DR Transport; Protein transport; Translocation; Transmembrane;
KW Inner membrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
SQ SEQUENCE 89 AA: 9664 MW: 4874F393EFFAD01B CPC64;
Query Match 3.0%. Score 6, DR 1, Length 89,
Best Local Similarity 100.0%; Prod. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 57 AKTEDA 62
DB 76 AKTEDA 81
Search completed: July 2, 2002, 16:19:07
Job time: 779 sec

Claim 11: SEQ ID NO 16877; 3071pp ; Sequence listing: English.

AAK54951 to AAK64762 encode the human immune/hematopoietic antigen (1) amino acid sequences given in AAK62176 to AAK61921. (1) have cytostatic activity, and can be used in gene therapy and vaccine production. (1) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (1) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (1) by expressing inactive proteins or to supplement the patient's own production of (1). Additionally, (1) polynucleotides may be used to produce the secreted (1), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (1) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic related diseases, especially cancers and cancer metastases of hematopoietic derived cells. AAK64703 to AAK87694 represent human immune/hematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54959 and AAK62169 represent sequences used in the exemplification of the present invention.

XX Sequence 77 AA;

Query Match 3.5%; Score 7; Bit 22; Length 77;
 Best local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 52 RLTSLSKK 58
 I I I I I I I I
 Db 27 RLTSLSKK 33

RESULT 5
 ABG15408
 ID ABG15408 standard; Protein: 98 AA.
 AC ABL15408;
 XX XX
 LT 18-FEB-2002 (first entry)
 XX XX
 DE Novel human diagnostic protein #15499.
 XX XX
 KW Human; chromosome mapping, gene mapping, gene therapy, forensic;
 KW Local supplement; medical imaging; diagnostic; genetic disorder;
 XX XX
 GS Homo sapiens.
 XX XX
 FN W0200175067-A2.
 XX XX
 PD 11-OCT-2001.
 XX XX
 PF 30 MAP 2001; 2001W-ns08631.
 XX XX
 PR 31-MAR-2000; 20000US-0540217.
 PR 23 AUG 2000; 20000US-0649167.
 XX XX
 PA (HYSE-) HYSEQ INC.
 XX XX
 PI Drmanac RT, Liu C, Tang YF;
 XX XX
 XX WPI, 2001 634362/73.
 DR N-PSDB; AAK579595.
 XX XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX XX
 PS Claim 20; SEQ ID NO 45767; 103pp; English.
 XX XX
 CC The invention relates to isolated polynucleotide (1) and
 CC polypeptide (11) sequences. (1) is useful as hybridisation probes,
 CC metastasis

```

PF 19-SEP-1997; 97WO-US10842.
XX 04-APR-1997; 97US-0042985.
PR 20-SEP-1996; 96US-0025296.
PR 03-APR-1997; 97US-0047611.
XX (TEXA ) UNIV TEXAS SYSTEM.
PA
XX Baer R, Bowcock AM;
PI WPI-1998-240417/20.
XX N-PSOH: AAV24135.
DP
XX DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which
PT as breast cancer antigen, BCRAL. Binding proteins are useful to
PT identify patient having or at risk of developing cancer
XX
XX Disclosure; page 287-288; 48pp; English.
XX
XX the sequence is that of a protein which can be used in the
CC preparation of the recombinant breast cancer antigen, BCRAL, binding
CC proteins BARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the
CC detection of a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid
CC sequence, specifically a wild type BARD1 composition for the detection
CC or purification of BCRAL, useful to identify a patient having, or at
CC risk of developing cancer. BARD1 can be used in the preparation of an
CC anti-BARD1 antibody, and in the detection and purification of a BCRAL
CC protein. BARD1, B123, BE2, BE14, BE31 or BE445 can be used in the
CC identification of a binding protein agonist or antagonist that alters
CC the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BCRAL or the
CC biological activity of the BCRAL-BARD1, B123, BE2, BE14, BE31 or
CC BE445 complex. The antibodies can be used to detect BARD1, B123, BE2, BE14,
CC BE31 or BE445, a specific anti-BARD1 antibody can be used to identify
CC a patient having or at risk of developing cancer.
XX
XX Sequence 515 AA;
SQ
Query Match 49.8%, Score 100, Db 19, Length 515,
Best Local Similarity 100.0%, Fract. No. 6.6%,
Matches 100; Conservative 0, Mismatches 0, Indels 0, Gaps 0;
OY 1 EYLLSALAAQKVVETLPHPTISYLYLQSLISQVHLHLKTLISEMSASQANIKLSLKKTL 60
Dh 125 EYLLSALAAQKVVETLPHPTISYLYLQSLISQVHLHLKTLISEMSASQANIKLSLKKTL 184
OY 61 ATTLPVLLPAIKTKYKQIEKNKNMNGPMSLQELHIG 100
Dh 195 ATTLPVLLPAIKTKYKQIEKNKNMNGPMSLQELHIG 244
RESULT 4
AAM89284
ID AAM89284 standard; protein; 77 AA.
AC AAM89284;
XX
XX (7-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen SEQ ID NO:16877.
XX
XX Human; immune; haematopoietic; immune/hematopoietic antigen; cancer,
KW cytostatic, gene therapy, vaccine, metastasis.
XX Homo sapiens.
XX WO200157182-A2
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR

```

```

PR 04 FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226661.
PR 22-AUG-2000; 2000US-0226668.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227609.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229504.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236402.
PR 02-OCT-2000; 2000US-0237037.

```

```

XX  Barry C. Bouqueret L. Chumakov L. Cohen-Akronine A;
XX  WPI: 2001-367042/48.
XX  N PSDB: AAF84099, AAF84010.
XX  New BAP28 polynucleotides and polypeptides overexpressed in prostate
XX  cancer cells for diagnosing prostate tumors, e.g. by hybridization or
XX  polymerase chain reaction assays.
XX  Claim 14: Page 297-304; 349pp; English.
XX  The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
XX  sequences and regulatory region located at the 3' and 5' ends of the
XX  BAP28 coding region. The BAP28 polypeptides can be expressed by standard
XX  recombinant methodology. BAP28 polynucleotides and polypeptides have been
XX  found to be over expressed in prostate tumour cells, therefore levels of
XX  BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
XX  reaction (PCR)) to diagnose patient suffering from or susceptible to
XX  prostate cancer. Antibodies specific for the BAP28 polypeptides are
XX  useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
XX  useful in genetic analysis. The present sequence represents a protein
XX  encoded by a first cDNA sequence of the BAP28 gene consisting of the
XX  exons 1 to 45.
XX  Sequence: 2144 AA;
XX
XX  Query Match: 59.5%; Score 269; BB 22; Length 2144;
XX  Best Local Similarity 100.0%; Pred. No. 7.7e-188;
XX  Matches 201; Conservative 9; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY 1 EYLLSALAAAGKVVETLPHFTSPYLEILSCVHLEITSEMGSASGANPLTSLKRTL 60
XX  Db 1754 EYLLSALAAAGKVVETLPHFTSPYLEILSCVHLEITSEMGSASGANPLTSLKRTL 1813
XX
XX  QY 61 ALIAPRVLLDAKKRYKQIEKNKNNHMCPEMSILOERIGXMKKFEILSHQSQTAFPTLE 120
XX  Db 1814 ALIAPRVLLDAKKRYKQIEKNKNNHMCPEMSILOERIGXMKKFEILSHQSQTAFPTLE 1873
XX
XX  QY 121 ALDFRAHSENGLEEVKGTENGICLAVMVVKLSEVTFRPLFFKLPDWAKTEDAPKRL 180
XX  Db 1874 ALDFRAHSENGLEEVKGTENGICLAVMVVKLSEVTFRPLFFKLPDWAKTEDAPKRL 1933
XX
XX  QY 181 LTFYNLAKWIAEKIKGLTFL 201
XX  Db 1934 LTFYNLAKWIAEKIKGLTFL 1954
XX
XX  RESULT 2
XX  AAB92729
XX  ID AAB92729 standard; Protein: 44 AA.
XX  AA
XX  AA
XX  26-JUN 2001 (first entry)
XX  Human protein sequence SEQ ID No.11159.
XX  Human: primer; detection: diagnosis; antisense therapy; gene therapy.
XX  Homo sapiens.
XX  EP1074617 A2.
XX  07 FEB 2001.
XX  28 JUL 2000; 2000JP-0116126.
XX  29 JUL 1999; 99JP-0249036.
XX  27-AUG-1999; 99JP-0300254.
XX  11 JAN 2000; 2000JP-0118776.
XX  02 MAY-2000; 2000JP-0183767.

```

```

XX  09-JUN-2000; 2000JP-0241899.
XX  (HELI-) HELIX RES INST.
XX  Ota T, Isozaki T, Nishikawa T, Hayashi K, Saio K, Yamamoto J;
XX  Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki I;
XX  WPI: 2001-318749/34.
XX  Primer sets for synthesizing polynucleotides, particularly the 5602
XX  full-length cDNAs defined in the specification, and for the detection
XX  and/or diagnosis of the abnormality of the proteins encoded by the
XX  full-length cDNAs.
XX  Claim 8, SEQ ID 11159, 2537pp + CD ROM, English.
XX  The present invention describes primer sets for synthesizing 5602
XX  full-length cDNAs defined in the specification. Where a primer set
XX  comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX  to the complementary strand of a polynucleotide which comprises one of
XX  the 5602 nucleotide sequences defined in the specification, where the
XX  oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX  of an oligonucleotide comprising a sequence complementary to the
XX  complementary strand of a polynucleotide which comprises a 5' end
XX  sequence and an oligonucleotide comprising a sequence complementary to a
XX  polynucleotide which comprises at least 15 nucleotides, where the
XX  oligonucleotide comprises at least 15 nucleotides and the combination of
XX  the 5' end sequence/3' end sequence is selected from those defined in
XX  the specification. The primer sets can be used in antisense therapy and
XX  in gene therapy. The primers are useful for synthesizing polynucleotides,
XX  particularly full-length cDNAs. The primers are also useful for the
XX  detection and/or diagnosis of the abnormality of the proteins encoded by
XX  the full-length cDNAs. The primers allow obtaining of the full length
XX  cDNAs easily without any specialised methods. AAH04166 to AAH14628 and
XX  AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to
XX  AAB95893 represent human amino acid sequences; and AAH13629 to AAH14632
XX  represent oligonucleotides, all of which are used in the exemplification
XX  of the present invention.
XX  Sequence 349 AA;
XX
XX  Query Match: 49.8%; Score 103; BB 22; Length 349;
XX  Best Local Similarity 100.0%; Pred. No. 4.5e-90;
XX  Matches 100; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX  QY 162 MKKFEILSHQSQTAFPTLEAFDFPFAJHSENGLEEVKGTENGICLAVMVVKLSEVTFR 161
XX  Db 60 MKKFEILSHQSQTAFPTLEAFDFPFAJHSENGLEEVKGTENGICLAVMVVKLSEVTFR 119
XX
XX  QY 162 LFFKLPDWAKTEDAPKDRLLITTYNLADCAEKIKGLTFL 201
XX  Db 120 LFFKLPDWAKTEDAPKDRLLITTYNLADCAEKIKGLTFL 159
XX
XX  RESULT 3
XX  AAW54099
XX  ID AAW54099 standard; Protein: 515 AA.
XX  AA
XX  AA
XX  28-SEP-1998 (first entry)
XX  Homo sapiens BAP28 sequence.
XX  HAPB1: ring protein; HPCAL: breast cancer; risk; diagnosis.
XX  Homo sapiens.
XX  OS
XX  W0961237 A2.
XX  26-MAR-1998.
XX

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:06:01 ; Search time 117.59 seconds
(without alignments)
189.862 Million cell updates/sec

Title: US 09 603 665 5_COPY_1754_1954

Perfect score: 201

Sequence: 1 EVYLISALAA:GKVVETPLP.....TFYNLADG:APKLC:PTLP 201

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 11107396 residues

Word size : 6

Total number of hits satisfying chosen parameters: 951

Minimum DB seq length: 0

Maximum DB seq length: 2040000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	200	99.5	2144	AAH85029	Protein encoded by
2	100	49.8	515	AAH85029	Human protein sequ
3	100	49.8	515	AAH85029	Human protein sequ
4	7	3.5	77	AAH85029	Human immune/hacma
5	7	3.5	77	AAH85029	Novel human diagno
6	7	3.5	165	AAH85029	H. vulgaris NIM1 h
7	7	3.5	195	AAH85029	Drosophila melanog
8	7	3.5	200	AAH85029	Fimeria merocrite
9	7	3.5	200	AAH85029	Drosophila melanog
10	7	3.5	216	AAH85029	Protein encoded by
11	7	3.5	216	AAH85029	Protein encoded by

12	7	3.5	216	14	AAH85029	Prod of the resb
13	7	3.5	245	22	AAH85029	Human protein sequ
14	7	3.5	254	22	AAH85029	Helicobacter pylor
15	7	3.5	289	22	AAH85029	Novel human diagno
16	7	3.5	292	22	AAH85029	Novel human diagno
17	7	3.5	302	22	AAH85029	Human C-protein co
18	7	3.5	311	21	AAH85029	Arabidopsis thalia
19	7	3.5	311	22	AAH85029	Human olfactory re
20	7	3.5	311	22	AAH85029	Human olfactory re
21	7	3.5	315	14	AAH85029	Surface antigen pr
22	7	3.5	320	22	AAH85029	Peptide #4239 enco
23	7	3.5	320	22	AAH85029	Peptide #3574 enco
24	7	3.5	321	22	AAH85029	Novel human diagno
25	7	3.5	323	22	AAH85029	Human olfactory re
26	7	3.5	323	22	AAH85029	Human olfactory re
27	7	3.5	323	22	AAH85029	Human olfactory re
28	7	3.5	327	21	AAH85029	Arabidopsis thalia
29	7	3.5	363	22	AAH85029	Novel human diagno
30	7	3.5	364	21	AAH85029	Arabidopsis thalia
31	7	3.5	371	22	AAH85029	Novel human diagno
32	7	3.5	417	22	AAH85029	Human protein sequ
33	7	3.5	431	20	AAH85029	Amino acid sequen
34	7	3.5	441	22	AAH85029	Human protein sequ
35	7	3.5	463	22	AAH85029	Human polypeptide
36	7	3.5	463	22	AAH85029	Alp-specific succi
37	7	3.5	463	22	AAH85029	Human protein sequ
38	7	3.5	463	22	AAH85029	DNA encoding human
39	7	3.5	485	22	AAH85029	Human polypeptide
40	7	3.5	521	22	AAH85029	Drosophila melanog
41	7	3.5	589	20	AAH85029	Drosophila melanog
42	7	3.5	593	22	AAH85029	Drosophila melanog
43	7	3.5	604	21	AAH85029	H. vulgaris NIM1 h
44	7	3.5	667	22	AAH85029	Enterococcus faeca
45	7	3.5	674	21	AAH85029	Human GPFX opp2245

ALIGNMENTS

RESULT 1
AAH85029
ID AAH85029 standard; Protein, 2144 AA.
AC AAH85029;
XX
XX
XX 06-AMG-2001 (first entry)
DE Protein encoded by HAP28 cDNA consisting of exons 1 to 45.
XX
XX HAP28, prostate; tumour, cancer, diagnostic, genetic analysis.
XX Homo sapiens
XX
XX Key location/Qualifiers
FH Misc difference 1694
FT /label= Ser or Asn
FT Misc-difference 1854
FT /label= Ala or Val
FT Misc-difference 1967
FT /label= Asp or Asn
FT Misc-difference 2017
FT /label= Gly or Glu
XX
XX Homo sapiens A3.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WC-1801183.
XX
XX 25-JUN-1999; 99NS-0141333.
XX 16-JAN-2000; 2000NS-0176880.
XX
XX (BEST) GENSMT.

DR PRINTS: PR00062; ALDOXETRIOLASE;
DR PROSITE: PS00062; ALDOXETRIOLASE_2; 1;
FT NON_TER 1
SQ SEQUENCE 252 AA: 28869 MW: 808305A5471D8A CRC64;

Query Match 4.58; Score 7; DB 11; Length 252;
Best Local Similarity 100.00; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VLALAEK 178
| | | | |
Eb 161 VLALAEK 167

Search completed: July 2, 2002, 16:18:32
Job time: 864 sec

```

AC 070473;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DR ALDEHYDE REDUCTASE (FRAGMENT)
OS Cricetus griseus (Chinese hamster)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
PP SEQUENCE FROM N.A.
RA Hyndman P.T., Flynn T.G.:
RT "Partial sequence of Chinese hamster aldehyde reductase.";
RI Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
DK EMBL; AF060820; AAC15760.1;
DR HSP; P14550; 2A1R.
DR InterPro: IPR001395; Aldo_ket_red.
DR Pfam: PF00248; Aldo_ket_red.
DR PRINTS: PR00069; ALDKREDTASE.
DR PROSITE: PS00798; ALDKREDTASE_1; 1.
DR PROSITE: PS00062; ALDKREDTASE_2; 1.
FT NON_TER 1
FT NON_TER 228
FT SEQUENCE 228 AA: 25450 MW: 90c5635FD810DA93 CPc64;

Query Match 3.5%; Score 7; DB 11; Length 228;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VLALAEK 178
DB 219 VLALAEK 225
|||||

RESULT 13
Q90WT1 PRELIMINARY; PRT; 232 AA.
AC Q90WT1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CRYSTALLIN B1 PROTEIN.
CN CRYSTALLIN B1.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
PP SEQUENCE FROM N.A.
RA Chen J.Y.;
RL Thesis (2001), Department of Institute of Zoology, Academia Sinica,
RL Taipei, Taiwan.
DR EMBL; AJ317957; CNC84899.1; 517HDM1A29AAC58R CPc64;
SQ SEQUENCE 232 AA: 26781 MW: 517HDM1A29AAC58R CPc64;

Query Match 3.5%; Score 7; DB 11; Length 232;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 YKIFLFD 90
DB 41 YKIFLFD 47
|||||

RESULT 14
Q9JPE8 PRELIMINARY; PRT; 238 AA.
AC Q9JPE8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)

```

```

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL. 24.8 KDA PROTEIN.
GN RTH41.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subphylum; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
PP SEQUENCE FROM N.A.
RC STRAIN=Z2491;
RX MEDLINE=20187481; PubMed=10722605;
RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,
RA Tinsley C.P.;
RT "Molecular and biological analysis of eight genetic islands that
RT distinguish neisseria meningitidis from the closely related pathogen
RT neisseria gonorrhoeae."
KO Infect. Immun. 68:2082-2095(2000).
DR EMBL; AJ391256; GAR71999.1;
KW Hypothetical protein.
SQ SEQUENCE 238 AA: 24810 MW: 52428E97A9A844R7 CPc64;

Query Match 3.5%; Score 7; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 DSSPKVR 165
DB 11 DSSPKVR 17
|||||

RESULT 15
Q9CT53 PRELIMINARY; PRT; 252 AA.
AC Q9CT53;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE 2610201A18RIK PROTEIN (FRAGMENT).
CN AKR1A1 OR 2510201A18RIK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
PP SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6J; TISSUE=EMRYO;
MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shindagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Hatalov S., Kasaiwa H.,
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Harsh G.,
RA Biele J., Hoffelli D., Hojunga N., Carninci P., de Renaldo M.F.,
RA Brownstein M.J., Hult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Bume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ping B., Kinkwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenck C., Soga I., Shibata Y., Storch K.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK011157; BAB27437.1;
DR HSP; P14550; 2A1R.
DR MGI; MGI:1929955; Akrlal.
DR InterPro: IPR001395; Aldo_ket_red.
DR Pfam: PF00248; Aldo_ket_red; 1.

```


Query Match: 4.5%; Score 7; DB 2; Length 98;
Best Local Similarity: 100.0%; Pred. NC: 32;
Matches: 7; Conservative: 0; Mismatches: 0; Indels:

27 29 EKLGLE 35
| | | | |
20 EKLGLE 26

RESULTS 9
0912V5
ID 0912V5
PRELIMINARY
PRT: 165 AA.

DT	01-OCT-2000	(TREMREL, 15, Created)
DT	01-OCT-2000	(TREMREL, 15, Last sequence update)
DT	01-OCT-2000	(TREMREL, 15, Last annotation update)

ON T20L15_180.
(OS Arabidopsis thaliana (Mouse-ear cross)).

X NCBI TaxID=3702;
Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots, Rosidae,
eurosids II; Brassicales; Brassicaceae; Arabidopsis.

RN
LLJ
RP
RA
RA
RA

SEQUENCE FROM N.A.
Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
De Clerk K., De Keyser A., Novt P., Ronze P., Van den Baele H.

RA Villalobos R., Gleason J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Budd S., Lemcke K., Mayer K.F.X.,
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RI Submitted (MAR) copy to the RMRU for
RA ED Arabidopsis sequencing project;
RP SEQUENCE FROM N.A.

DR EMBL; AF162451; CAB827611;
KW Hypothetical protein.
SQ SEQUENCE 165 AA; 16092 MW.
2140B73BA998BA2B CRR64.

Query Match	3.5%	Score 7;	DB 10;	Length
Best Local Similarity	100.0%	Prod. No.	51;	

QY	176	AEKLEN	18
DB	30	AEKLEN	46

RESULT 10
Q96Z15
10 Q96Z15
PRELIMINARY: PRE: 167 AA.

DT	01-DEC-2001	(TREMBlrel. 19, Created)
DT	01-DEC-2001	(TREMBlrel. 19, Last sequence update)
DT	01-DEC-2001	(TREMBlrel. 19, Last annotation update)

GN ST2014.

```

      ARCADE ('ARCADE') AS ARCADE,
      NBI_TaxID-111955,
      FN [1]
    
```

RE SEQUENCED FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed-11572479;

DR EMBL; AP000988; BAB67111.1; -
KW Hypothetical protein; Complete
SQ SEQUENCE 167 AA; 18937 MW; 9C68D37EC65FH719 CRC64;

Query Match 3.5%; Score 7; DB 17; Length 167;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels

Qy 30 K1.KGLFT 36
Db 72 KLKGLFT 78

RESULT 11
Q9D0P3
ID Q9D0P3
PRELIMINARY:
PRT: 214 AA

AC	Q9DUP37	01-JUN-2001	(TREMBlrel. 17, Created)
DT		01-JUN-2001	(TREMBlrel. 17, Last sequence update)
DT		01-JUN-2001	(TREMBlrel. 17, Last annotation update)

DE 2610201A18RIK PROTEIN.
CN AKR1A1 OR 2610201A18RIK.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
NCBI_TaxID=10090;
111
PN

RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-EM3RYO;
RX MEDLINE-21085660; PubMed-11217851;

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii
RA Aizawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka
RA

RA Kadota K., Matsuda H.A., Ashbu
RA Fleischmann W., Gaasterland I.

RA Schriml L.M., Staubli F., Suzuki R.
RA Sakai K., Okido T., Furuno M., Aono

RA Brownstein M.J., Bult C., Plet
RA Gustincich S., Hill D., Holman

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto
RA Sasaki H., Sato K., Schoenbach C., Soya T., Shibata Y.,
RA

RA Hayashizaki Y.;
RT "Functional annot
KI. Nature 409:685-68

DR MGD; M
DR InterP
DR Plam;

GN GYRF
OS Flexibacter sancti
OC Bacteria; CFB group; Flexibacter group; Flexibacter.
OX NCBI_TaxID=1004;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IP015057;
RA Suzuki M., Yamauchi K.;
RT "Phylogenetic analysis and taxonomic study of marine Cytophaga like
bacteria. Proposal of Haerentibaculum gen. nov. with Haerentibaculum
maritimum comb. nov. and Haerentibaculum ovoliticus comb. nov., and
two new species.";
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RL Submitter: ATP DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC -!- CATALYTIC ACTIVITY: ATP DEPENDENT BREAKAGE, PASSAGE AND REJOINING
OF DOUBLE-STRANDED DNA.
CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC EMBL: AB032582; BAB13320.1; -;
DR HSSP: P06982; 1A16.
DR InterPro: IPR001241; DNA_topoisomII.
DR InterPro: IPR003594; HATPase_c.
DR InterPro: IPR002936; Toprim.
DR Pfam: PF00204; DNA_topoisomII; 1.
DR Pfam: PF02518; HATPase_c; 1.
DR Pfam: PF01751; Toprim; 1.
DR PRINTS: PR00418; TPI2FAMILY.
DR ProDom: PD000616; DNA_topoisomII; 1.
DR SMART: SM00433; TOP2C; 1.
DR POSITE: PS00177; TOPOISOMERASE_II; 1.
KW ATP-binding; Isomerase; Topoisomerase.
FT NON-TER 1
FT NON-TER 481
SQ SEQUENCE 481 AA; 53517 MW; 1F3AB57936BC10F2 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 SENDPEKC 72
DB 316 SENDPEKC 323

RESULT 6
ID O9A014 PRELIMINARY; PRT; 481 AA.
AC O9A014;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE DNA GYRASE B SUBUNIT (FRAGMENT).
GN GYRF.
OS Chitinophaga pinensis.
OC Bacteria; CFB group; Flexibacter group; Chitinophaga.
OX NCBI_TaxID=79329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 2588;
RA Suzuki M., Takadara T., Harayama S., Yamamoto S.;
RT "Diversity of marine Cytophaga-like bacteria: Phylogenetic analysis
using gyrB sequences and their carotenoids profiles.";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
OF DOUBLE-STRANDED DNA.
CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC EMBL: AB048186; BAB33153.1; -;
DR HSSP: P06982; 1A16.
DR InterPro: IPR001241; DNA_topoisomII
DR InterPro: IPR003594; HATPase_c.
DR InterPro: IPR002936; Toprim.
DR Pfam: PF00204; DNA_topoisomII; 1.
DR Pfam: PF02518; HATPase_c; 1.
DR Pfam: PF01751; Toprim; 1.

DR PRINTS: PR00418; TPI2FAMILY.
DR ProDom: PD000616; DNA_topoisomII; 1.
DR SMART: SM00433; TOP2C; 1.
DR POSITE: PS00177; TOPOISOMERASE_II; 1.
KW ATP-binding; Isomerase; Topoisomerase.
FT NON-TER 1
FT NON-TER 481
SQ SEQUENCE 481 AA; 53681 MW; E8452B58870427F2 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 SENDPEKC 72
DB 316 SENDPEKC 323

RESULT 7
ID O44537 PRELIMINARY; PRT; 72 AA.
AC O44537;
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE ORF 11.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria, gamma subdivision, Pseudomonadaceae;
OC Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89123097; PubMed-2644218;
RA Jacobson M.R., Brigle K.E., Bennett L.F., Setterquist R.A.,
PA Wilson M.S., Cash V.L., Beynon J., Newton W.E., Dean D.R.;
RT "Physical and genetic map of the major nif gene cluster from
Azotobacter vinelandii.";
RL J. Bacteriol. 171:1017-1027 (1989).
DR EMBL: M20568; AAA64721.1; -;
SQ SEQUENCE 72 AA; 8004 MW; HCRCA2D8E3035F8B CRC64;

Query Match 3.5%; Score 7; DB 2; Length 72;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 ALAEKLR 180
DB 13 ALAEKLR 24

RESULT 8
ID O31026 PRELIMINARY; PRT; 98 AA.
AC O31026;
DT 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DT 01-JAN-1998 (TREMREL. 05, Last annotation update)
DE HYDROTHERMAL 11.2 KDA PROTEIN.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria, gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=569B;
RA Kaewrakon P., Manning P.A.;
RL Submitted (SEP 1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF025662; AAB81983.1; -;
KW Hypothetical protein.
SQ SEQUENCE 98 AA; 11215 MW; 615AC18896D4F84 CRC64;

01-MAR-2001 (TREMREL. 16, Last sequence update)
 01-DEC-2001 (TREMREL. 19, Last annotation update)
 MHC CLASS II ALPHA CHAIN.
 HPIIALPHA.
 OS Calman crocodilus (Spectacled calman) (Calman sclerops).
 OC Fukuyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Crocodylia; Alligatorinae; Calman.
 NC NCHI_TaxID=8499;
 RN [1]
 RP SEQUENCE FROM N.A.
 KA Voldby J., Vilved L., Due M., Gronlund J., Holmskov U., Teisner B.,
 KA Salomonsen J., Brusaard K., Skjold K.:
 KI "Cloning, sequence and genomic structure of MHC class II antigens from
 KI the spectacled calman Calman crocodilus."
 KI Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF256650; AAF9982.1;
 DR HSSP: P01908; 25EB.
 DR InterPro: IPR004597; 4_C1.
 DR InterPro: IPR004006; 4_MHC.
 DR InterPro: IPR001009; MHC_II_alpha.
 DR Pfam: PF00047; 1q; 1.
 DR Pfam: PF00994; MHC_II_alpha; 1.
 DR SMART: SM00407; 1q; 1.
 DR PROSITE: PS00290; 1C_MHC; UNKNOWN_1.
 SQ SEQUENCE 253 AA: 28550 MW: 1945FA5D9060E4325 CRC64;

Query Match 4.0%; Score 8; DB 7; Length 253;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ACTIVLAL 175

DB 13 ALITVLAL 20

RESULT 3

Q9FAX2
 ID Q9FAX2 PRELIMINARY: PRT: 481 AA.
 AC Q9FAX2;
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE DNA GYRASE B SUBUNIT (FRAGMENT).
 GN GYRB.
 OS Flexibacter japonensis.
 OC Bacteria; CF8 group; Flexibacter group; Flexibacter.
 NC NCHI_TaxID=104662;
 RN [1]
 RP SEQUENCE FROM N.A.
 KA Suzuki M., Yamaguchi K.:
 KI "Phylogenetic analysis and taxonomic study of marine Cytophaga like
 KI bacteria. Proposal of Haerentibaculum gen. nov., with Haerentibaculum
 KI maritimum comb. nov., and Haerentibaculum ovoides comb. nov., and
 KI two new species."
 KI Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 DR EMBL: AB042579; BAB13317.1;
 DR HSSP: P06982; 1AJ6.
 DR InterPro: IPR001241; DNA_topoisomII.
 DR InterPro: IPR004594; HATPase_c.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF00204; DNA_topoisomII; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR PRODOM: P0000616; DNA_topoisomII; 1.
 DR SMART: SM00433; TOP2c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
 KW ATP-binding; Isomerase; Topoisomerase.
 FT NON_TER 1
 FT NON_TER 481
 SQ SEQUENCE 481 AA: 53735 MW: 1E4FD8E3E4EEFA30 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 SENDPEKC 72

DB 316 SENDPEKC 323

RESULT 5

Q9FAW9
 ID Q9FAW9 PRELIMINARY: PRT: 481 AA.
 AC Q9FAW9;
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE DNA GYRASE B SUBUNIT (FRAGMENT).

FT NON_TER 1
 FT NON_TER 481
 SQ SEQUENCE 481 AA: 53556 MW: 8707FEA30EDA70B1 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 SENDPEKC 72

DB 316 SENDPEKC 323

RESULT 4

Q9FAX0
 ID Q9FAX0 PRELIMINARY: PRT: 481 AA.
 AC Q9FAX0;
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE DNA GYRASE B SUBUNIT (FRAGMENT).
 GN GYRB.
 OS Flexibacter filiformis.
 OC Bacteria; CF8 group; Flexibacter group; Flexibacter.
 NC NCHI_TaxID=104663;
 RN [1]
 RP SEQUENCE FROM N.A.
 KA Suzuki M., Yamaguchi K.:
 KI "Phylogenetic analysis and taxonomic study of marine Cytophaga like
 KI bacteria. Proposal of Haerentibaculum gen. nov., with Haerentibaculum
 KI maritimum comb. nov., and Haerentibaculum ovoides comb. nov., and
 KI two new species."
 KI Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA
 DR EMBL: AB032581; BAB13319.1;
 DR HSSP: P06982; 1AJ6.
 DR InterPro: IPR001241; DNA_topoisomII.
 DR InterPro: IPR004594; HATPase_c.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF00204; DNA_topoisomII; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR PRODOM: P0000616; DNA_topoisomII; 1.
 DR SMART: SM00433; TOP2c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
 KW ATP-binding; Isomerase; Topoisomerase.
 FT NON_TER 1
 FT NON_TER 481
 SQ SEQUENCE 481 AA: 53735 MW: 1E4FD8E3E4EEFA30 CRC64;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: July 2, 2002, 16:18:31, Search time: 96.19 Seconds
(without alignments)
361.493 Million cell updates/sec

Title: US-09-603-665-5_COPY_1917_211
 Perfect score: 201
 Sequence: 1 FKLFWAKTERDAPKDRLLTF.....NYVLLPESIPFLAELMEDE 201

Scoring table:	01.100	
Gapop	60.0	Gapopt 60.0

562222 seqs, 172994929 residues
searched.

Word size : 6

Model	Number of hits satisfying chosen parameters	1509
Model 1	1	1
Model 2	1	1
Model 3	1	1
Model 4	1	1
Model 5	1	1
Model 6	1	1
Model 7	1	1
Model 8	1	1
Model 9	1	1
Model 10	1	1
Model 11	1	1
Model 12	1	1
Model 13	1	1
Model 14	1	1
Model 15	1	1
Model 16	1	1
Model 17	1	1
Model 18	1	1
Model 19	1	1
Model 20	1	1
Model 21	1	1
Model 22	1	1
Model 23	1	1
Model 24	1	1
Model 25	1	1
Model 26	1	1
Model 27	1	1
Model 28	1	1
Model 29	1	1
Model 30	1	1
Model 31	1	1
Model 32	1	1
Model 33	1	1
Model 34	1	1
Model 35	1	1
Model 36	1	1
Model 37	1	1
Model 38	1	1
Model 39	1	1
Model 40	1	1
Model 41	1	1
Model 42	1	1
Model 43	1	1
Model 44	1	1
Model 45	1	1
Model 46	1	1
Model 47	1	1
Model 48	1	1
Model 49	1	1
Model 50	1	1
Model 51	1	1
Model 52	1	1
Model 53	1	1
Model 54	1	1
Model 55	1	1
Model 56	1	1
Model 57	1	1
Model 58	1	1
Model 59	1	1
Model 60	1	1
Model 61	1	1
Model 62	1	1
Model 63	1	1
Model 64	1	1
Model 65	1	1
Model 66	1	1
Model 67	1	1
Model 68	1	1
Model 69	1	1
Model 70	1	1
Model 71	1	1
Model 72	1	1
Model 73	1	1
Model 74	1	1
Model 75	1	1
Model 76	1	1
Model 77	1	1
Model 78	1	1
Model 79	1	1
Model 80	1	1
Model 81	1	1
Model 82	1	1
Model 83	1	1
Model 84	1	1
Model 85	1	1
Model 86	1	1
Model 87	1	1
Model 88	1	1
Model 89	1	1
Model 90	1	1
Model 91	1	1
Model 92	1	1
Model 93	1	1
Model 94	1	1
Model 95	1	1
Model 96	1	1
Model 97	1	1
Model 98	1	1
Model 99	1	1
Model 100	1	1

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

post-processing: listing first 45 summaries

```
Database :
SPPREMRL_10: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_page: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_ivirus: *
16: sp_bacterioph: *
17: sp_archaeo: *
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	100	49.8	34.9	4	Q96RS5	Q96RS5	homo sapien
2	8	4.0	25.3	7	Q9GJN1	Q9GJN1	caiman croc
3	6	4.0	48.1	2	Q7IAX2	Q7IAX2	flexibacter
4	8	4.0	48.1	2	Q9FAX0	Q9FAX0	flexibacter
5	8	4.0	48.1	2	Q9FAW9	Q9FAW9	flexibacter
6	8	4.0	48.1	2	Q9AQ14	Q9AQ14	chitinophag
7	7	3.5	72	2	Q44537	Q44537	azotobacter
8	7	3.5	98	2	O31026	O31026	vibrio chol
9	7	3.5	165	10	Q91ZV5	Q91ZV5	arabidopsis
10	7	3.5	167	17	Q96Z15	Q96Z15	sulfolobus
11	7	3.5	214	11	Q9D0P3	Q9D0P3	mus musculus
12	7	3.5	228	11	O70473	O70473	cricetulus
13	7	3.5	232	13	Q90WF1	Q90WF1	brachydanio
14	7	3.5	238	2	Q9JHP8	Q9JHP8	neisseria m
15	7	3.5	252	11	Q9CT53	Q9CT53	mus musculus
16	7	3.5	252	11	Q9CQ15	Q9CQ15	mus musculus

17	7	3.5	265	16	Q98HM1	Q98hm1 rhizobium 1
18	7	3.5	285	6	Q9N045	Q9n045 macaca fasc
19	7	3.5	296	16	Q9AAU6	Q9aau6 caulobacter
20	7	3.5	305	2	Q935V5	Q935v5 streptomyce
21	7	3.5	317	5	P91020	p91020 caenorhabdi
22	7	3.5	325	11	Q9J116	Q9j116 mus musculus
23	7	3.5	325	11	Q9D017	Q9d017 mus musculus
24	7	3.5	325	11	Q9C014	Q9c017 mus musculus
25	7	3.5	325	11	Q9D012	Q9d012 mus musculus
26	7	3.5	325	11	Q9C018	Q9c018 mus musculus
27	7	3.5	326	16	Q55684	Q55684 synchocyst
28	7	3.5	332	10	Q9FK92	Q9fk92 arabidopsis
29	7	3.5	333	13	Q9TG17	Q9d017 brachydanio
30	7	3.5	336	2	Q9Z361	Q9z361 staphylococ
31	7	3.5	337	2	Q9R968	Q9r968 staphylococ
32	7	3.5	337	3	P95738	P95738 staphylococ
33	7	3.5	337	5	Q45326	Q45326 caenorhabdi
34	7	3.5	342	4	Q96G28	Q96g28 homo sapien
35	7	3.5	343	11	Q9CWQ4	Q9cwq4 mus musculus
36	7	3.5	343	11	Q99JR9	Q99jr9 mus musculus
37	7	3.5	353	17	Q28890	Q28890 archaebactob
38	7	3.5	359	5	Q21008	Q21008 caenorhabdi
39	7	3.5	374	5	Q9V313	Q9v313 drosophila
40	7	3.5	374	10	Q9C502	Q9c502 arabidopsis
41	7	3.5	391	2	Q9S2B3	Q9s2b3 streptomyce
42	7	3.5	393	16	Q9REQ0	Q9req0 bacillus ha
43	7	3.5	397	3	Q9C1T3	Q9c1t3 yarrowia 11
44	7	3.5	400	16	Q9KSH9	Q9ksb9 vibrio chol
45	7	3.5	401	17	Q97513	Q97513 sulfobobus

ALIGNMENTS

	RESULT	I
ID	Q96E55	PPELIMINARY; PRT; 349 AA.
AC	Q96E55:	
CD	01-DEC-2001 (TRIMBUREL_19, Created)	
DT	01-DEC-2001 (TRIMBUREL_19, Last sequence update.)	
DF	01-DEC-2001 (TRIMBUREL_19, Last annotation update)	
DE	SIMILAR TO HYPOTHETICAL PROTEIN FLJ10359.	
DS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Ethelicia; Primates, Catarrhini, Hominoidea; Homo.	
OX	NCBI_TaxID=9606;	
NH	111	
RP	SEQUENCE FROM N.A.	
RPC	TISSUE-OVARY, AND ADENOCARCINOMA;	
RL	Straussberg R.;	
RA	Submitted (Jul-2001) to the EMBL/GenBank/DDBJ databases.	
DR	EMBL: BC011983; AALH1983.1 ;	
DN	SEQUENCE 349 AA: 39921 MW: 3A3359597FF7F7079FR CRG64;	

RA Hosouchi I., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimpo S., Takouchi C., Wada I., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RF entire genome and assignment of potential protein-coding regions.";
 RL DNA RES. 3:109-136(1996).
 CC -! FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE
 CC NUCLEOTIDE BIOSYNTHESIS.
 CC -! CATALYTIC ACTIVITY: GTP + IMP + L-aspartate -> GDP + phosphate +
 CC adenylosuccinate.
 CC -! PATHWAY: FIRST COMMITTED STEP IN AMP BIOSYNTHESIS.
 CC -! SIMILARITY: BELONGS TO THE ADENYLOSUCCINATE SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL: D90405; BAA1743.1; ALT INIT.
 DR BSSP: P12283; LADE.
 DR InterPro: IPRO0114; Adenylosucc_Synt.
 DR Pfam: PF00709; Adenylosucc_Synt; 1.
 DR ProDom: PD01188; Adenylosucc_Synt; 1.
 DR PROSITE: PS00513; ADENYLOSUCIN_SYN_2; 1.
 DR PROSITE: PS01266; ADENYLOSUCIN_SYN_1; 1.
 KW Purine biosynthesis; ligase; GTP-binding; Complete proteome.
 FI NP_HUNG 14 19 GIP (POTENTIAL).
 FI ACT_SITE 140 140 HY SIMILARITY.
 FI ACT_SITE 147 147 HY SIMILARITY.
 SQ SEQUENCE 444 AA: 48792 MW: F06900ECF241A4BF CRG64;

Query Match 3.58; Score 7; DH 1; Length 444;
 Best Local Similarity 100.0%; Pred.No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 FLAEIME 199
 |||
 DB 400 FLAEIME 406

Search completed: July 2, 2002, 16:19:08
 Job time: 780 sec

```

KW Apoptosis; Zinc-finger; Repeat.
FT REPEAT 4 70 BIR 1.
FT REPEAT 90 157 BIR 2.
FT REPEAT 193 283 CARD.
FT DOMAIN 311 346 RING-TYDE.
FT ZN_FING 358 AA, 40977 MW; EB2268FA9A6190A4 CPG64;
SQ SEQUENCE 358 AA, 40977 MW; EB2268FA9A6190A4 CPG64;

Query Match
Best Local Similarity 3.5%; Score 7; DB 1; Length 358;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KDLTLTF 20
DB 5 KDLTLTF 11

RESULT 13
YQJH_BAC50
ID YQJH_BAC50 STANDARD; PRT: 414 AA.
AC P54545;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 47.0 kDa protein in qinQ-ansr intergenic region.
GN YQJH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
SQ SEQUENCE FROM N.A.
RP STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE IMPE/MUCE/SAMP FAMILY
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: D84432; BAA12614.1; -
CC EMBL: Z99116; CAB14319.1; -
CC Subtilist; BG11737; YQJH.
CC InterPro: IPR001126; UMG0C.
CC Pfam: PF00817; IMS; 1.
CC Hypothetical protein; Complete proteome
KW SEQUENCE 414 AA, 47020 MW; 0824C62080B8448+ CPG64;
SQ SEQUENCE 414 AA, 47020 MW; 0824C62080B8448+ CPG64;

Query Match
Best Local Similarity 3.5%; Score 7; DB 1; Length 414;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 AEKLGK 34
DB 193 AEKLGK 199

RESULT 14
YCL4_PYRHO
ID YCL4_PYRHO STANDARD; PRT: 441 AA.
AC 058969;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter extracellular binding protein PH1214
DE precursor.

GN PH1214 OR PHBK040.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
PC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Haba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Utsuka K., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushida N., Oouchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Fob F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RT DNA Res. 5:55-76(1998).
RI -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
RI SYSTEM PH1214/15/16.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: AP000005; BAA30314.1; -
CC InterPro: IPR000567; SHP_bac_1.
CC Pfam: PF01547; SHP_bacterial_1; 1.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC PROSITE: PS01037; SBI_BACTERIAL_1; FALSE NEG.
CC Hypothetical protein; Transport; Membrane; Lipoprotein; Signal;
CC Complete proteome.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 441 EXTRACELLULAR ABC TRANSPORTER
CC LIPID 24 24 EXTRACELLULAR BINDING PROTEIN PH1214.
CC SEQUENCE 441 AA, 48896 MW; 558D963D816CA7EE CPG64;

Query Match
Best Local Similarity 3.5%; Score 7; DB 1; Length 441;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 AEKLEKN 182
DB 191 AEKLEKN 197

RESULT 15
PURA_SYNY3
ID PURA_SYNY3 STANDARD; PRT: 444 AA.
AC P73290;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)
DE (ADSS) (AMP/ase).
DE PURA OR SH1823.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneke T., Sato S., Kofani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,

```

or send an email to license@isb-sib.ch).

CC EMBL: D10854; BAA01627.1; -
 DR HSSP: P14550; ZAIR.
 DR InterPro: IPR001495; Aldo_ket_red.
 DR Pfam: PF00248; aldo_ket_red_1.
 DR PRINTS: P00069; ALDOKETRED.
 DR PROSITE: PS00062; ALDOKETO-REDUCTASE_2; 1.
 DR PROSITE: PS00063; ALDOKETO-REDUCTASE_4; 1.
 DR PROSITE: PS00798; ALDOKETO-REDUCTASE_1; 1.
 KW oxidoreductase; NADP; Acetylation.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 FT A_T_SITE 112 112 HYDROXY-INDOL DONOR (BY SIMILARITY).
 SQ SEQUENCE 124 AA: 16374 MW: 43855A5F0F9D5F2 CRC64:

Query Match 3.5% Score 7; DB 1; Length 124;

Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VIATAFK 178
 DQ 234 VIATAFK 240

RESULT 11

ID ADHA_RHIME STANDARD; PRT: 340 AA.
 AC 031186;
 DT 30 MAY 2000 (Rel. 39, Created)
 DI 01 MAY 2000 (Rel. 39, Last sequence update)
 DE Alcohol dehydrogenase (EC 1.1.1.1).
 GN ADHA OR RA0704 OR SMA1296.
 OS Rhizobium meliloti (SinoRhizobium meliloti).
 OG Plasmid pSymA (megaplasmid 1).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 RX Rhizobiaceae; SinoRhizobium.
 XX NCBI_TaxID:382;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021.
 RA MEDLINE:98424558; PubMed:9659380;
 RA Willis L.H., Walker G.C.;
 RT Identification of the Rhizobium meliloti alcohol dehydrogenase gene
 (adha) and heterologous expression in Alcaligenes eutrophus.*;
 RL Biochim. Biophys. Acta 1984;197:203(1998).
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021.
 RA MEDLINE:2196509; PubMed:1181432;
 RA Harcourt M.J., Fisher R.F., Jones L., Komp C., Abola A.P.,
 RA Barclay Hubler F., Bowser L., Capella D., Galibert F., Gaudy J.,
 RA Gurjal M., Hong A., Huizar F., Hyman P.W., Kahn D., Kahn M.L.,
 RA Kalman S., Keating P.H., Palm C., Peck M.C., Szczepki R., Wells D.H.,
 RA Yeh K.C., Davis R.W., Federlespi N.A., Long S.R.;
 RT Nucleotide sequence and predicted functions of the entire
 SinoRhizobium meliloti pSymA megaplasmid.*;
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 CC -/- COFACTOR: ZINC (BY SIMILARITY).
 CC -/- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 FAMILY. HIGH, WITH OTHER HACIERIAL ADH'S.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

CC EMBL: AF031940; AAB87463.1; -
 DR EMBL: AF007258; AAK65362.1; -
 DR HSSP: P00325; IOPH.
 DR InterPro: IPR002328; Adh_znf_family.
 DR InterPro: IPR002085; Adh_znf_family.
 DR Pfam: PF00107; adh_zinc_1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW oxidoreductase; zinc; NAD; Plasmid; Complete proteome.
 FT METAL 40 40 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 63 63 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 340 AA: 36235 MW: C1A04B5FEE5C95E1 CRC64:

Query Match 3.5% Score 7; DB 1; Length 340;

Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 LALAEKL 179
 DQ 203 LALAEKL 209

RESULT 12

ID PIAP_PIG STANDARD; PRT: 358 AA.
 AC 062640;
 DT 15-DEC-1998 (Rel. 37, Created)
 DI 15-DEC-1998 (Rel. 37, Last sequence update)
 DI 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative inhibitor of apoptosis.
 GN PIAP.
 OS Sus scrofa (pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 UX NCBI_TaxID:9823;
 RN 111
 RP SEQUENCE FROM N.A.
 RC 11SSUP-Aorta.
 RX MEDLINE:98162622; PubMed:9501011;
 RA Stehlik C., de Martin R., Finkler R.R., Lipp J.;
 RT "Cytokine induced expression of porcine inhibitor of apoptosis
 protein (Iap) family member is regulated by NF kappa B.*;
 RL Biochem. Biophys. Res. Commun. 243:827-832(1998).
 CC -/- SIMILARITY: BELONGS TO THE IAP FAMILY.
 CC -/- SIMILARITY: CONTAINS 2 HIR REPEATS.
 CC -/- SIMILARITY: CONTAINS 1 CARD DOMAIN.
 CC -/- SIMILARITY: CONTAINS 1 RING TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

CC EMBL: U75142; AAC39171.1; -
 DR HSSP: O13490; IOPH.
 DR InterPro: IPR001370; BIR.
 DR InterPro: IPR001315; CARD.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF00653; HIR; 2.
 DR Pfam: PF00619; CARD; 1.
 DR Pfam: PF00097; zf-C3HC4; 1.
 DR SMART: SM00238; BIR; 2.
 DR SMART: SM00114; CARD; 1.
 DR SMART: SM00184; PIR; 1.
 DR PROSITE: PS01282; BIR_REPEAT_1; 2.
 DR PROSITE: PS50143; BIR_REPEAT_2; 2.
 DR PROSITE: PS50209; CARD; 1.
 DR PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
 DR PROSITE: PS50089; ZF_RING_2; 1.

```

CC EMBL: AF000001; HAA29183.1; -.
DR InterPro: IPR000445; MoaA_NLH_PqqE.
DR Pfam: PF01444; MoaA_NLH_PqqE; 1.
DR PROSITE: PS01305; MOA_NLH_PQQE; 1.
KW Molybdenum cofactor biosynthesis; Iron-sulfur; Complete proteome.
DR METAL: 27 27 IRON-SULFUR (POTENTIAL).
FT METAL 31 31 IRON-SULFUR (POTENTIAL).
FT METAL 34 34 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 316 AA; 60.0% MW; A2BFE1747A7F58A; J0644;

Query Match 3.5%; Score 7; DB 1; Length 316;
Best local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 LAELKE 181
DB 105 LAELKE 111

RESULT 9
ALDX_HUMAN STANDARD; PRT; 324 AA.
ID ALDX_HUMAN
DC P14550;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAY-2002 (Rel. 41, Last annotation update)
DE Alcohol dehydrogenase [NADP+] (EC 1.1.1.2) (Aldehyde reductase)
GN AKR1A1 OR ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SOURCE:
RC TISSUE=livor;
RX MEDLINE=87289801; PubMed=3615425;
RA Wermuth B., Omar A., Forster A., di Francesco C., Wolf M.,
RA von Wartburg J.-P., Bullock B., Gabbay K.H.;
RA "Primary structure of aldehyde reductase from human liver.";
RA Prog Clin Biol Res 232:247-267(1987)
RN 12;
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=89255461; PubMed=2494333;
RA Bohren K.M., Bullock B., Wermuth B., Gabbay K.H.;
RA "The aldo keto reductase superfamily cDNAs and deduced amino acid
RA sequences of human aldehyde and aldose reductases.";
RA J. Biol. Chem. 264:9547-9551(1989).
RN 13;
RP SEQUENCE FROM N.A
RA Fujii J.;
RA Submitted (DEC 1997) to the EMBL/GenBank/DBJ databases.
RN 14;
RP SEQUENCE FROM N.A.
RX MEDLINE=94417674; PubMed=10486210;
PA Barski O.A., Gabbay K.H., Bohren K.M.;
PA "Characterization of the human aldehyde reductase gene and
PA promoter.";
PA Genomics 60:188-198(1999).
RN 15;
RP X-RAY CRYSTALLOGRAPHY (2.48 ANGSTROMS)
PA El-Kabbani O., Green N.G., Lin G., Carson M., Narayana S.V.L.,
PA Moore K.M., Flynn T.G., DeLucas L.J.;
PA "Structures of human and porcine aldehyde reductase: an enzyme
PA implicated in diabetic complications.";
PA Acta Crystallogr. D 50:856-868(1994).
CC 1- CATALYTIC ACTIVITY: An alcohol + NADP(+) -> an aldehyde + NADPH.
CC 1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/

```

```

CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to licenses@isb.sib.ch).
CC -----
DR EMBL: J04794; AAA51711.1; -.
DR EMBL: AF036683; AAR92469.1; -.
DR EMBL: AF036680; AAB92369.1; JOINED.
DR EMBL: AF036681; AAR92465.1; JOINED.
DR EMBL: AF036682; AAB92369.1; JOINED.
DR EMBL: AF112485; AAF01260.1; -.
DR EMBL: AF112484; AAF01260.1; JOINED.
DR PIR: A33851; A33851.
DR PDB: 2ALP; 20-UN-96.
DR SWISS-2DPAGE; P14550; HUMAN.
DR MIM: 103830; -.
DR InterPro: IPR001395; Aldo_ket_red.
DR Pfam: PF00248; aldo_ket_red; 1.
DR PRINTS: PR00069; ALDKETREDTASE.
DR PROSITE: PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE: PS00063; ALDOKETO_REDUCTASE_3; 1.
DR PROSITE: PS00798; ALDOKETO_REDUCTASE_1; 1.
KW Oxidoreductase, NADP, Acetylation, 3D structure.
FT INIT_MET 0 0
FT MOD_RES 1 1 ACETYLATION.
FT ACT_SITE 11? 11? HYDROGEN-BOND DONOR (BY SIMILARITY)
SQ SEQUENCE 324 AA; 36.44% MW; AAG50065A940F1F CPC64;

Query Match 3.5%; Score 7; DB 1; Length 324;
Best local Similarity 100.0%; Pred No 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VLALAEK 178
DB 233 VLALAEK 239

RESULT 10
ALDX_RAT STANDARD; PRT; 324 AA.
ID ALDX_RAT
DC P51635;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alcohol dehydrogenase [NADP+] (EC 1.1.1.2) (Aldehyde reductase)
GN AKR1A1 OR ALR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11;
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Kidney, and Liver;
RX MEDLINE=9427440; PubMed=6500767;
PA Takahashi M., Fujii J., Teshima T., Suzuki K., Shiba T., Taniguchi N.;
PA "Identity of a major 3-deoxyglucose-reducing enzyme with aldehyde
PA reductase in rat liver established by amino acid sequencing and cDNA
PA expression.";
PA Gene 127:249-253(1993).
RN 12;
RP FUNCTION: CATALYZES THE NADPH-DEPENDENT REDUCTION OF 3-
RP DEHYXYGLUCOSE (3-DG).
CC 1- CATALYTIC ACTIVITY: An alcohol + NADP(+) -> an aldehyde + NADPH.
CC 1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/

```



```

CC Helicobacter.
CC NCBI_TaxID=85963;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=99120557; PubMed=924682;
CC Alm R.A., Ling L.-S.L., Moir D.I., King H.L., Brown E.D., Doig P.C.,
CC Smith D.R., Noonan B., deJonge H.L., Carmel G.,
CC Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Iwas C.,
CC Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
CC Trust T.J.;
CC *Genomic sequence comparison of two unrelated isolates of the human
CC gastric pathogen Helicobacter pylori.*;
CC Nature 397:171-183(1999);
CC -!- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-xetionate 8-phosphate +
CC phosphate -> phosphoenolpyruvate + D-arabinose 5-phosphate + H(2O).
CC -!- PATHWAY: FIRST STEP IN LIPOLYSACCHARIDE BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE KDSA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF001440; AAD05587.1;
CC InterPro: IPR001785; DAHP_synth_1;
CC Pfam: PF00793; DAHP_synth_1;
CC Lipopolysaccharide biosynthesis: Lyase; Complete proteome.
CC SEQUENCE: 276 AA; 30374 MW; EBF7DB79FD48D3H8 CR664;

Query Match 4.5%; Score 7; DB 1; Length 276;
Best Local Similarity 00.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LKTRSS 161
DB 148 LKTRSS 154
|||||

RESULT 7
OTC_ARCFU
ID OTC_ARCFU STANDARD: PRT: 407 AA.
AC 029013;
DT 15-JUL-1998 (Rel. 36, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCase).
GN ARGF OR AF1255.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae.
CX Archaeoglobus.
XX NCBI_TaxID=2234;
XX [1]
XX SEQUENCE FROM N.A.
XX STRAIN=Vr-16 / DSM 4404 / ATCC 49589;
XX MEDLINE=98049443; PubMed=9489475;
XX Klenk H.P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
XX Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
XX Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
XX Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
XX Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
XX Peterson S., Reich C.L., McNeil L.K., Hedges J.H., Glodek A., Zhou L.,
XX Overberg K., Gwagye J.D., Weidman J.F., McDonald L., Utterback T.,
XX Cotton M.B., Springs J., Artchak P., Kaine B.P., Sykes S.M.,
XX Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
XX Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
XX Venter J.C.;
XX *The complete genome sequence of the hyperthermophilic, sulphate
XX reducing archaeon Archaeoglobus fulgidus.*;

```

```

RL Nature 390:364-370(1997);
CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-ornithine -> phosphate
CC + L-citrulline.
CC -!- PATHWAY: ARGinine BIOSYNTHESIS (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE AICASES/OTCASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001017; AAB89987.1;
CC HSSP: Q51742; IALS.
CC TIGR: AF1255;
CC
CC InterPro: IPR002029; Carbmyltransf_asor.
CC Pfam: PF00185; OTCace; 1;
CC Pfam: PF02729; OTCace; 1;
CC PRINTS: PR00100; OTCASE
CC PROSITE: PS00867; CARBAMYLTRANSFEPASE; FALSE_NEG.
CC Transfease: Arginine biosynthesis; Complete proteome.
CC SEQUENCE: 307 AA; 34913 MW; DB043855A40E8H5D CR664;

Query Match 3.5%; Score 7; DB 1; Length 307;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 LAEKLE 181
DB 21 LAEKLE 27
|||||

RESULT 8
MOAA_PYRHO
ID MOAA_PYRHO STANDARD: PRT: 316 AA.
AC 057854;
DT 30-MAY-2000 (Rel. 39, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative molybdopterin cofactor synthesis protein A.
GN MOAA OR PH0114.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
CX NCBI_TaxID=53953;
XX [1]
XX SOURCE: FROM N.A.
XX STRAIN=OT3;
XX MEDLINE=98344147; PubMed=9679194;
XX Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
XX Yamamoto S., Sekine M., Haba S.-i., Kosugi H., Hosoyama A., Naito Y.,
XX Sakai M., Ogura K., Otsuwa P., Nakazawa H., Takamiya M., Ohfuku Y.,
XX Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Onuchi A.,
XX Aoki K.-I., Yoshizawa T., Nakamura Y., Osada F., Horikoshi K.,
XX Masuchi Y., Shizuya H., Kikuchi H.;
XX *Complete sequence and gene organization of the genome of a hyper
XX thermophilic archaeobacterium, Pyrococcus horikoshii OT3.*;
XX DNA Res. 5:555-76(1998).
XX -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF MOLYBDOPTERIN PRECURSORS
XX Z FROM GUANOSINE (BY SIMILARITY).
XX -!- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS (BY SIMILARITY).
XX -!- SIMILARITY: BELONGS TO THE MOAA / NIFB / POOE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

Wed Jul 3 09:02:32 2002

```

QY 181 ENYIVLL 187
DB 128 ENYIVLL 134

RESULT 4
RL9_MYCGE STANDARD; PRT: 150 AA.
AC 147339; 1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L9.
GN RPL19 OR P19 OR MG093
OS Mycoplasma genitalium
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;

SEQUENCE FROM N.A.
MEDLINE=85130985; PubMed=3882694;
Wold W.S.M., Cladaras C., Deutscher S.L., Kapoor O.S.;
"the 19-kDa glycoprotein coded by region E3 of adenovirus.
Purification, characterization, and structural analysis.";
J. Biol. Chem. 260:2424-2431(1985).
[3]
COMPLETE GENOME.
MEDLINE=92087470; PubMed 1727603;
Chroboczek J., Hiebert F., Jacrot H.,
"the sequence of the genome of adenovirus type 5 and its comparison
with the genome of adenovirus type 2.";
Virology 186:280-285(1992).
-1- FUNCTION: FORMS A COMPLEX WITH THE HEAVY CHAIN OF THE CLASS I
TRANSPLANTATION ANTIGENS DURING AN ADENOVIRUS INFECTION. THERMOLY
REDUCING CLASS I EXPRESSION ON THE CELL SURFACE.
-1- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic
reticulum.
-1- SIMILARITY: TO OTHER ADENOVIRUSES GP19K.
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL: X33260; NOT_ANNOTATED_CDS.
EMBL: M33002; CAA26783.1;
EMBL: M12406; AAA42492.1;
PIR: A03822; ERA0A5
P1R: A22515; A22515.
Early protein; transmembrane; Glycoprotein; Signal.
POTENTIAL.
CHAIN 1 18
EARLY E3 18.5 KDA GLYCOPROTEIN.
LUMENAL (POTENTIAL).
TRANSMEM 125 145
POTENTIAL.
DOMAIN 146 160
CYTOPLASMIC (POTENTIAL).
DOMAIN 30 30
N-LINKED (GLCNAC...) (POTENTIAL).
CARBOHYD 79 79
SEQUENCE 160 AA, 18502 MW, B1F0D2AC4D5340E0 CRC64;

Query Match 3.5%; Score 7; DH 1; Length 160;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 LITVLAL 175
DB 132 LITVLAL 138
|||||

RESULT 6
KDSA_HELPJ STANDARD; PRT: 276 AA.
ID KDSA_HELPJ
AC Q92N55;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2-dehydro-3-deoxythiophosphate aldolase (EC 4.1.2.16) (Phospho-2-
dehydro-3-deoxycyclonate aldolase) (3-deoxy-D-manno-octulosonic acid
8-phosphate synthetase) (KDO-8-phosphate synthetase) (KDO 8-P
synthase).
GN KDSA OR JHP0003.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

QY 174 ALAEKIK 180
DB 63 ALAEKIK 69
|||||

RESULT 5
F3GL_ADE05 STANDARD; PRT: 160 AA.
ID F3GL_ADE05
AC P04494;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Early E3 18.5 kDa glycoprotein precursor (GP19K).
OS Human adenovirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28285;
RN [1]

```

```

FI  VAR:ANI 2017 2017 /F110-VAR.010941.
FI  E 1.0
FI  /FT.0 VAL.010042
SQ  SEQUENCE 2144 AA: 242455 MW: 5666L6EE76DC9B7 CRC64:

Query Match
Best local Similarity 100.0%; DB 1; Length 2144;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ALMPPLVDQLENLGLVEERKFEQVETVTHLPCTAQFSVAMADGDSWKPLNYQILLKTRDSS 161
DB 2019 ALMPPLVDQLENLGLVEERKFEQVETVTHLPCTAQFSVAMADGDSWKPLNYQILLKTRDSS 2077
QY 162 PKVRFALITVLALAKFKENYIVLLPSIPFLAELMELE 201
DB 2078 PKVRFALITVLALAKFKENYIVLLPSIPFLAELMELE 2117

RESULT 2
RP28_MACFA
ID RP28_MACFA STANDARD: PRC: 958 AA:
AC Q96M44:
DI 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein BA28 (Fragment).
GN BAP28.
OS Macaca fascicularis (Orang eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cerropithecoidea;
OC Cercopithecoinae; Macaca.
OX NCBI_TaxID=9541:
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Suzuki N., Hida M., Kusuda J., Tanaka K., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sudano S., Hashimoto K.,
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RT Submitted (OCT 2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BAP21 FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HEAT REPEAT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB049842; BAB16728.1; ALT_INIT.
DR InterPro: IPR000457; HEAT_repeat.
DR PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.
FT NON_TER 1
FT REPEAT 920 956 HEAT.
SQ SEQUENCE 958 AA: 108644 MW: 3DBD95C3623CFB41 CRC64:

Query Match
Best local Similarity 100.0%; DB 1; Length 958;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLFWAKTDAFKORLLTFYNLADQIAEKLGKLTFLAGHLVKPFDNLT 50
DB 731 PKLFWAKTDAFKORLLTFYNLADQIAEKLGKLTFLAGHLVKPFDNLT 780

RESULT 4
CYIS_RAT
ID CYIS_RAT
AC P19113:
STANDARD: PRC: 141 AA:

```

```

DI 01-NOV-1990 (Rel. 16, Created)
DI 01-FEB-1996 (Rel. 33, Last sequence update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cystatin S precursor (LM protein).
GN CST4 OR CYSS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Submandibular gland;
RX MEDLINE=92165056; PubMed=1537554;
RA Cox J.L., Shaw P.A.;
RT "Structure, organization and regulation of a rat cysteine proteinase
RT inhibitor-encoding gene."
RL Gene 110:175-180(1992).
RN [2]
RP SEQUENCE OF 10-141 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Submandibular gland;
RX MEDLINE=89053983; PubMed=3263967;
RA Shaw P.A., Cox J.L., Barka T., Naito Y.;
RT "Cloning and sequencing of cDNA encoding a rat salivary cysteine
RT proteinase inhibitor inducible by beta-adrenergic agonists."
RL J. Biol. Chem. 263:18133-18137(1988).
RN [3]
RP SEQUENCE OF 28-132.
RC TISSUE=Submandibular gland;
RX MEDLINE=89334379; PubMed=2757366;
RA Bedi G.S.;
RT "Amino acid sequence of an inducible cysteine proteinase inhibitor
RT (cystatin) from submandibular glands of isoproterenol treated rats."
RL Arch. Biochem. Biophys. 274:245-253(1989).
CC -!- FUNCTION: THIS PROTEIN STRONGLY INHIBITS PAPAIN & FIGIN. PARTIALLY
CC INHIBITS STEM BROMELAIN & BOVINE CATHEPSIN C, BUT DOES NOT INHIBIT
CC PORCINE CATHEPSIN B OR CLOSTRIPAIN. PAPAIN IS INHIBITED
CC NONCOMPETITIVELY.
CC -!- TISSUE SPECIFICITY: FOUND IN SALIVA, TEARS, URINE, AND SEMINAL
CC FLUID.
CC -!- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M75281; AAA41068.1;
DR FMBL: J04206; AAB59703.1;
DR PIR: A31891; A31891.
DR PIR: S05252; S05252.
DR HSSP: P01038; ICEW.
DR InterPro: IPR000010; Cystatin.
DR InterPro: IPR003243; Cystatin_C_M.
DR Pfam: PF00031; cystatin; 1.
DR ProDom: PD001231; Cystatin_C_M; 1.
DR SMART: SM00043; CY; 1.
DR PROSITE: PS00287; CYSTATIN; 1.
KW Thiol protease inhibitor; Signal.
FT SIGNAL 1 27
FT CHAIN 28 141 CYSTATIN S.
FT ACT_SITE 32 32 REACTIVE SITE.
FT SITE 76 80 SECONDARY AREA OF CONTACT.
FT DISULFID 94 104 BY SIMILARITY.
FT DISULFID 118 138 BY SIMILARITY.
FT CONFLICT 114 115 EH -> OE (IN REF. 4).
SQ SEQUENCE 141 AA: 15949 MW: D7632905541CR266 CRC64:

Query Match
Best local Similarity 100.0%; DB 1; Length 141;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

genCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:14:07 : Search time 30.19 seconds
(without alignments)
257.788 Million cell updates/sec

Title: US-09-603-665-5_COPY_1917_2117
Perfect score: 201
Sequence: 1 PFLDMWAKTADAPKRLITFNYIVLLPESIPFLAELMEDE 201

Scoring table: OLICO

Gapop 60.0 , Gapext 60.0
Searched: 105224 seqs, 38719550 residues

Word size : 6

Total number of hits satisfying chosen parameters. 359

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Note: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	49.8	2144	1 BP28_HUMAN	Q9H583 homo sapien
2	50	24.9	958	1 P6226 MACFA	Q9q44 macaca fasc
3	7	3.5	141	1 CITS_RAT	P9313 rattus norv
4	7	3.5	150	1 RL9_RICSE	P47339 mycoplasma
5	7	3.5	160	1 E3G_ADE05	P04494 human adeno
6	7	3.5	276	1 KDS_HELPJ	Q9zn55 helicobacte
7	7	3.5	307	1 QTC_ARCFU	Q29013 archaeoglob
8	7	3.5	316	1 M0AL_PYRHO	O57854 pyrococcus
9	7	3.5	324	1 ALDZ_HUMAN	P14550 homo sapien
10	7	3.5	324	1 ALDZ_RAT	P51635 rattus norv
11	7	3.5	340	1 ADHP_RHIME	Q31146 rhizobium m
12	7	3.5	358	1 PIAL_PIG	O62640 sus scrofa
13	7	3.5	414	1 YQJF_BACSU	P94545 bacillus su
14	7	3.5	441	1 YC14_PYRHO	O58969 pyrococcus
15	7	3.5	444	1 PHRA_SYNY3	P73290 synechocyst
16	7	3.5	463	1 VDHX_CHICK	Q90578 gallus gall
17	7	3.5	502	1 Y190_HELPJ	P56117 helicobacte
18	7	3.5	556	1 PHR_TARAU	P34205 carassius a
19	7	3.5	885	1 CHS3_EXODH	P30602 exophiala d
20	7	3.5	911	1 CHSG_ASPEU	P34267 aspergillus
21	7	3.5	916	1 CHSB_EKENT	Q00757 emericella
22	7	3.5	1045	1 IMD2_YEAST	P12684 saccharomyc
23	7	3.5	1121	1 DDX8_AFAIH	Q38953 arabidopsis
24	7	3.5	1122	1 ADPL_MYCGA	Q49379 mycoplasma
25	7	3.5	1612	1 ATCA_YEAST	G12675 saccharomyc
26	7	3.5	1630	1 SPJ8_YEATH	Q9c844 arabidopsis
27	6	3.0	68	1 YE02_HREN	P44177 haemophilus
28	6	3.0	73	1 RL7_STRAU	P48860 staphylococ
29	6	3.0	84	1 EX7S_HAEHL	P43914 haemophilus
30	6	3.0	87	1 GLXK_HAEHL	P45242 haemophilus
31	6	3.0	89	1 TAVA_ECOLI	O65938 escherichia
32	6	3.0	96	1 REV_SIVAM	P36339 simian immu
33	6	3.0	101	1 REV_SIVGB	P22379 simian immu

RESULT 1

ID	BP28_HUMAN	STANDARD;	PRT;	2144 AA.
AC	Q9H583; Q9NM23;			
DI	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Protein BAP28.			
GN	BAP28.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. AND VARIANTS S 1694, A 1854, D 1967 AND G 2017.			
RA	Rouquieret E., Chumakov I., Barry C., Cohen-Akenine A.;			
RT	"A novel BAP28 gene and protein."			
RL	Patent number WO0100669, 04-JAN-2001.			
RN	[2]			
RP	SEQUENCE OF 1534-2144 FROM N.A.			
RA	Cobley V.;			
RL	Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.			
RN	[3]			
PP	SEQUENCE OF 1777-2144 FROM N.A.			
PA	Isoqai T., Ota T., Hayashi K., Suiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiraiqi A., Sudo H.,			
RA	Wakatsuma M., Hosori T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
PA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,			
RA	Ninomiya K., Iwayanagi T.;			
RT	"NEDO human cDNA sequencing project."			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.			
CC	1. SIMILARITY. BELONGS TO THE BAP28 FAMILY.			
CC	1-1. SIMILARITY. CONTAINS 1 HEAT REPEAT.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb.scrib.ch)			
CC	or send an email to license@isb.scrib.ch .			
CC	EMBL; AX067150; CAC26776.1; ..			
DR	EMBL; ALI36105; CAC15948.1; ..			
DR	EMBL; AK001221; BAA91564.1; ALT_INIT.			
DR	InterPro; IPRO000357; HEAT_REPEAT.			
DR	PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.			
KW	POLYMERPHISM.			
FT	REPEATS 2105 2142			
FT	VARIANT 1694 1854			
FT	HEAT N -> S.			
FT	/FTID=VAR_010939.			
FT	VARIANT 1854 1854			
FT	V -> A.			
FT	/FTID=VAR_010940.			
FT	VARIANT 1967 1967			
FT	N -> D.			

ALIGNMENTS

Wed Jul 3 09:02:31 2002

Query Match 3.5%; Score 7; DB 2; Length 325;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VLALAEK 178
| | | | |
Db 234 VLALAEK 240

Search Completed. July 3, 2002, 16:15:41
Job time: 758 sec

C:Species: Archaeoglobus sulfidus
C:Date: 05 Dec 1997 #sequence_revision 05 Dec 1997 #text_change 18-Jun-1999
C:Accession: F69406
R:Klink, R.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.C.; Gill, S.; Kirkness, E.F.;
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 464-470, 1997
A:Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Arriach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
A:Reference number: A69250; M011:9604944
A:Accession: F69406
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1407 <K1K>
A:Cross-references: GB:AF001017; GB:AF000742; NID:q2689340; PIDD:AA89987.1; PIDD:q264932
C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase
F:2-400/domain; aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 3.5% Score 7; DB 2; Length 307;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 LAEKLKE 181
IIIIII
DB 21 LAEKLKE 27

RESULT 12
H71241
molybdopterin biosynthesis protein moaA - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 24 Oct 2000
C:Accession: H71241
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yanamoto, S.; Seki
M.; Ohtoku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Koshida, N.; Ouchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; M011:94044137
A:Accession: H71241
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-416 <KAW>
A:Cross-references: GB:AF000001; NID:q1246128; PIDD:BA229183.1; PIDD:q3256500
A:Experimental source: Strain O1
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank
C:Genes: PH0114
C:Superfamily: Escherichia coli molybdopterin biosynthesis protein moaA
C:Keywords: molybdopterin biosynthesis

Query Match 3.5% Score 7; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 LAEKLKE 181
IIIIII
DB 105 LAEKLKE 111

RESULT 13
L25526
hypothetical protein C0708.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: L25526
R:Nelson, J.
A:Submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid C0708.
A:Reference number: Z25045
A:Accession: L25526

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-317 <NEI>
A:Cross-references: EMBL:080844; PIDD:AA837476.1; GSPDH:GN00028; GESP:C0708.6
A:Experimental source: strain Bristol N2; clone C0708
C:Genetics:
A:Gene: GESP:C0708.6
A:Map position: X
A:Exons: 81/3; 218/2; 303/2
C:Superfamily: aldehyde reductase

Query Match 3.5% Score 7; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VIAlAEK 178
IIIIII
DB 234 VIAlAEK 240

RESULT 14
JN0629
alcohol dehydrogenase (EC 1.1.1.1) - rat
N:Alternate names: aldehyde reductase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Apr-2000 #sequence_revision 20 Apr 2000 #text_change 19 Jun 2000
C:Accession: JN0629; P05532
R:Takahashi, M.; Fujii, J.; Teshima, T.; Suzuki, K.; Shiba, T.; Taniguchi, N.
Gene 127, 249-253, 1993
A:Title: Identity of a major 3-deoxyglucosone-reducing enzyme with aldehyde reductase
A:Reference number: JN0629; M01:9327340
A:Accession: JN0629
A:Molecule type: mRNA
A:Residues: 1-325 <TAK>
A:Cross-references: GB:D10854; NID:q399659; PIDD:BA01627.1; PIDD:q399660
A:Accession: P05532
A:Molecule type: protein
A:Residues: 14 23; 35-45; 69-77; 81-91; 135-145; 241 249; 295-305; 309-318 <1AA>
C:Comment: This enzyme catalyzes the NADPH dependent reduction of 3-deoxyglucosone to
ns.
C:Superfamily: aldehyde reductase
C:Keywords: alcohol metabolism; NADP; oxidoreductase
F:263/Binding site: NADP (Lys) <status predicted>

Query Match 3.5% Score 7; DB 1; Length 425;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VIAlAEK 178
IIIIII
DB 234 VIAlAEK 240

RESULT 15
A33851
alcohol dehydrogenase (NADP+) (EC 1.1.1.2) - human
C:Species: Homo sapiens (man)
C:Date: 23-Mar-1990 #sequence_revision 23 Mar 1990 #text_change 11-Jun-1999
C:Accession: A33851
R:Bohren, K.M.; Bullock, B.; Wermuth, B.; Gabbay, K.H.
J. Biol. Chem. 264, 9547-9551, 1989
A:Title: The aldo-keto reductase superfamily. cDNAs and deduced amino acid sequences
A:Reference number: A33851; M01:89255461
A:Accession: A33851
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-325 <BOH>
A:Cross-references: GB:J04794; NID:q178480; PIDD:AAA51711.1; PIDD:q178481
C:Superfamily: aldehyde reductase
C:Keywords: alcohol metabolism; NADP; oxidoreductase

Wed Jul 3 09:02:31 2002

A:Accession: T48212
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-165 <REV>
 A:Cross-references: EMBL:AJ162351
 A:Experimental source: cultivar Columbia; BAC clone T20L15
 C:Genetics:
 A:Map position: 5
 A:Introns: 99/3; 121/3
 A:Note: T20L15.180

Query Match 3.5% Score 7; DB 2; Length 165;

Best Local Similarity 100.0%; Pred. No. 24;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ABEKLEN 182

DB 30 ABEKLEN 36

RESULT 7

JC7706

beta crystallin B1 protein - zebra fish

C:Species: Brachydanio rerio (zebra fish)

C:Date: 04 Nov-2001 #sequence_revision 00-Nov-2001 #text_change 27-Nov-2001

C:Accession: JC7706

R:Chen, J.Y.; Chang, B.F.; Chen, Y.H.; Lin, C.J.F.; Wu, J.L.; Kuo, C.M.

Biochem. Biophys. Res. Commun. 285, 105-110, 2001

A:Title: Molecular cloning, developmental expression, and hormonal regulation of zebrafish

A:Reference number: JC7706; MIM:21331298; PMID:11437479

A:Accession: JC7706

A:Molecule type: mRNA

A:Residues: 1-232 <CH>

A:Cross-references: GB:AJ317957

C:Comment: This protein is involved in regulation by growth factors.

C:Superfamily: beta-crystallin

F:20-25/Region: active proline- and alanine-rich motif #status predicted

Query Match 3.5% Score 7; DB 2; Length 232;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 YKIFLED 90

DB 41 YKIFLED 47

RESULT 8

K71985

3-deoxy-d mann-octulosonic acid 8-phosphate synthase - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12 Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: E71985

R:Alm, R.A.; Ling, L.S.L.; Moir, D.F.; King, B.L.; Brown, E.D.; Deig, P.C.; Smith, D.K.;

Ives, C.; Gibson, R.; McBryde, D.; Mills, S.; Jiang, Q.; Taylor, D.R.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MIM:99120557

A:Accession: E71985

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-276 <ARN>

A:Cross-references: GB:AE001440; GB:AE001439; NID:q4154501; FIDN:AND05587.1; FID:q415450

A:Experimental source: strain J99

C:Genetics:

A:Gene: kdsA

C:Superfamily: phospho-2-dehydro 3-deoxyoctonate aldolase

Query Match 3.5% Score 7; DB 2; Length 276;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LKTRDSS 161

DB 148 LKTRDSS 154

RESULT 9

AG3305

23S ribosomal RNA methyltransferase (EC 2.1.1.1) [imported] Brucella melitensis (st

C:Species: Brucella melitensis

C:Date: 01 Feb 2002 #sequence_revision 01 Feb 2002 #text_change 01 Feb 2002

C:Accession: AG3305

Kishinevskiy, V.G.; Kaputai, V.; Pedkar, P.; Patra, G.; Mulier, C.; Los, T.; Ivanov

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit

A:Reference number: AD3252; PMID:11756688

A:Accession: AG3305

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-276 <KOR>

A:Cross-references: GB:AE008917; FIDN:AA51610.1; FID:q17982446; GSPDR:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMF10429

A:Map position: 1

C:Keywords: methyltransferase

Query Match 3.5% Score 7; DB 2; Length 276;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 LENRGG 117

DB 67 LENRGG 73

RESULT 10

AB7311

hypothetical protein CC0498 [imported] Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20 Apr 2001 #text_change 20-Apr-2001

C:Accession: AB7311

R:Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, P.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; KO

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus

A:Reference number: AB7249; MIM:21173698; PMID:11259647

A:Accession: AB7311

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-296 <STG>

A:Cross-references: GB:AE005673; NID:q13421677; FIDN:AAK22485.1; GSPDR:GN00148

C:Genetics:

A:Gene: CC0498

Query Match 3.5% Score 7; DB 2; Length 296;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LEFTFAG 40

DB 240 LEFTFAG 246

RESULT 11

F69406

ornithine carbamoyltransferase (argF) homolog - Archaeoglobus fulgidus

GenCore version 4.5
Copyright (c) 1993 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:15:40 ; Search time 57.00 seconds
(without alignments)

338,307 Million cell updates/sec

Title: US-09-603-665-5_COPY_1917_2117

Perfect score: 201

Sequence: 1 FKLEWDKATEDAPKPLTF NVIVLLPESIPFLAELMEDE 201

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 28438 seqs, 9608934 residues

Word size : 6

Total number of hits satisfying chosen parameters: 1011

Minimum DB seq length: 0

Maximum DB seq length: 20000000

Post processing: listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	ID	Description
1	7	3.5	98	2 S26914	Ig heavy chain V r
2	7	3.5	141	2 JG1470	cystatin s precursor
3	7	3.5	150	2 C64210	ribosomal protein
4	7	3.5	151	2 AC3385	retrovirus-related
5	7	3.5	160	1 ERADA5	early E3 18.5K gly
6	7	3.5	165	2 T48212	hypothetical prote
7	7	3.5	232	2 JC7706	beta crystallin H1
8	7	3.5	276	2 JF1985	3-deoxy-d-manno-oc
9	7	3.5	276	2 AC3305	23S ribosomal RNA
10	7	3.5	296	2 A87311	hypothetical prote
11	7	3.5	307	2 F60406	ornithine carbamoy
12	7	3.5	316	2 H71231	molybdopterin bios
13	7	3.5	317	2 T25526	hypothetical prote
14	7	3.5	325	1 JN0629	alcohol dehydrogen
15	7	3.5	325	2 A33851	alcohol dehydrogen
16	7	3.5	326	2 S76360	hypothetical prote
17	7	3.5	337	2 T20644	hypothetical prote
18	7	3.5	340	2 H95349	alcohol dehydrogen
19	7	3.5	353	2 I69422	P420-nonreducing h
20	7	3.5	358	2 JC5264	apoptosis inhibito
21	7	3.5	359	2 T22950	hypothetical prote
22	7	3.5	372	2 AE3184	alcohol dehydrogen
23	7	3.5	374	2 B86457	unknown protein. 1
24	7	3.5	391	2 T36739	hypothetical prote
25	7	3.5	393	2 G83749	methionine gamma l
26	7	3.5	397	2 JC7575	acetoacetyl-CoA re
27	7	3.5	400	2 C82213	conserved hypothet
28	7	3.5	414	1 H69063	DNA-damage repair
29	7	3.5	415	2 S77471	adenylosuccinate s

ALIGNMENTS

RESULT 1

S26914
Ig heavy chain V region (DP-2) - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 22 Nov 1992 #sequence_revision 10 Nov 1995 #text_change 21 Jan 2000

C:Accession: S26914

J. Tomlinson, I. M., Walter, G., Marks, J. D., Lilwelyn, M. B., Winter, G.

J. Mol. Biol. 227, 776-798, 1992

A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups o

A:Reference number: S26885; MUID:93021117

A:Accession: S26914

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-98 <DOM>

A:Cross-references: EMBL 212304; NID 33863; FIEN CAA78174.1; FID 332864

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotrimer; immunoglobulin

F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 3.5% Score 7; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 KFOPRVT 126

|||||||

Db 63 KFOERTV 69

RESULT 2

JQ1470

cystatin S precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 17 Jul 1992 #sequence_revision 17 Jul 1992 #text_change 16 Jul 1999

C:Accession: JQ1470; A31891; S05252; S15132

R. Cox, J. L. Shaw, P. A.

Gene 110, 175-180, 1992

A:Title: Structure, organization and regulation of a rat cysteine proteinase inhibito

A:Reference number: JQ1470; MUID:92165056

A:Accession: JQ1470

A:Molecule type: DNA

A:Residues: 1-141 <COX>

A:Cross-references: GB M75281; NID 9294537; FIEN AAA41068.1; FID 3294548

P. Shaw, P. A.; Cox, J. L.; Harka, T.; Naito, Y

J. Biol. Chem. 263, 18133-18137, 1988

A:Title: Cloning and sequencing of cDNA encoding a rat salivary cysteine proteinase 1

A:Reference number: A31891; MUID:89053983

A:Accession: A31891

A:Molecule type: mRNA

A:Residues: 171STDYLYISFEHTLSC', 27 141 -SHA-

P. Redl, G. S

Arch Biochem Biophys 273, 245-253, 1989

adenylosuccinate s
hypothetical prote
hypothetical prote
D-alanyl-D-alanine
probable cytoplasm
ABC transporter pe
vitamin D3 hydroxy
conserved hypothet
phosphatase - goldf
2',3'-cyclic-nucle
mobilisation prote
kelch-like protein
iron (II) transpor
polyadenylate-bind
probable membrane-
hypothetical prote



us-09-603-665-5_copy_1917_2117.olig6.ra1

Wed Jul 3 09:02:29 2002

Db 10 AEKLE 15

Search completed. July 3, 2002, 16:04:01
Job time: 58 sec

COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb Storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/928,930A
FILING DATE: 19920812
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5506218man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DCKET NUMBER: R060 206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 848-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-928 930A 3

Query Match 3.0% Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 AEKLE 181
DB 10 AEKLE 15

RESULT 14
US-08 288 568 3
Sequence 3, Application US/08288568
Patent No. 5587366
GENERAL INFORMATION:
APPLICANT: Parker, Thomas S.; Levine, Daniel M.;
APPLICANT: Rubin, Albert L.; Gordon, Bruce R.; Saal, Stuart D.
TITLE OF INVENTION: Methods Useful in Endotoxin Based
TITLE OF INVENTION: Prophylaxis and Therapy
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb
MEDIUM TYPE: Storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08288,568
FILING DATE: 10-AUGUST-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07453
FILING DATE: 9-AUGUST-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5506218man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DCKET NUMBER: R060 211
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-288-568-3

Query Match 3.0% Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 AEKLE 181
DB 10 AEKLE 15

RESULT 15
US-08-487-461-3
Sequence 3, Application US/08187461
Patent No. 5587366
GENERAL INFORMATION:
APPLICANT: Parker, Thomas S.; Levine, Daniel M.;
APPLICANT: Rubin, Albert L.; Gordon, Bruce R.; Saal, Stuart D.
TITLE OF INVENTION: Methods Useful in Endotoxin Based
TITLE OF INVENTION: Prophylaxis and Therapy
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb
MEDIUM TYPE: Storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08487,461
FILING DATE: June 7, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07453
FILING DATE: 9-AUGUST-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587366man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DCKET NUMBER: R060 211.2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-487-461-3

Query Match 3.0% Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 AEKLE 181
DB 10 AEKLE 15

```

: FILING DATE: 19901012
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Gamson, Edward P.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312 616-5400
: TELEFAX: 312 616-5460
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1045 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-596-467-6

Query Match 3.5%; Score 7; DB 1; Length 1045;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 HLKPKFA 47
Db 10 HLKPKFA 16

RESULT 11
US-07-934-374-6
: Sequence 6, Application US/07934374
: Patent No. 5349126
: GENERAL INFORMATION:
: APPLICANT: Chappell, J.
: APPLICANT: Saunders, Court A.
: APPLICANT: Cucular, R.
: APPLICANT: Wolf, Fred R.
: TITLE OF INVENTION: Process and Composition for Increasing
: TITLE OF INVENTION: Sterol Accumulation in Higher Plants
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dressler, Goldsmith, Shore, Sutker & Milnamow
: STREET: 180 N. Statson St.
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION NUMBER: US/07/934,374
: FILING DATE: 19920814
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Gamson, Edward P.
: REGISTRATION NUMBER: 29,381
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312 616-5400
: TELEFAX: 312 616-5460
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1045 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-934-374-6

Query Match 3.5%; Score 7; DB 1; Length 1045;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 HLKPKFA 47
Db 10 HLKPKFA 16

FILING DATE: 19901012
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312 616-5400
TELEFAX: 312 616-5460
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-596-467-6

Query Match 3.5%; Score 7; DB 1; Length 1045;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 HLKPKFA 47
Db 10 HLKPKFA 16

RESULT 12
US-07-783-861C-6
: Sequence 6, Application US/07783861C
: Patent No. 5460949
: GENERAL INFORMATION:
: APPLICANT: Saunders, Court A.
: APPLICANT: Wolf, Fred R.
: APPLICANT: Mukharil, Indrani
: TITLE OF INVENTION: A Method and Composition for Increasing
: TITLE OF INVENTION: The Accumulation of Squalene and Specific Sterols in
: TITLE OF INVENTION: Yeast
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Amoco Corp, Patents and Licensing Dept.
: STREET: 200 East Randolph St.
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60680-0703
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION NUMBER: US/07/783,861C
: FILING DATE: 19911028
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/613,380
: FILING DATE: 15-NOV-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Galloway, No. 5460949vall B.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312 856-7180
: TELEFAX: 312 856-4972
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1045 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-783-861C-6

Query Match 3.5%; Score 7; DB 1; Length 1045;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 HLKPKFA 47
Db 10 HLKPKFA 16

RESULT 13
US-07-928-930A-3
: Sequence 3, Application US/07928930A
: Patent No. 5348422
: GENERAL INFORMATION:
: APPLICANT: Levine, Daniel M.; Parker, Thomas S.;
: APPLICANT: Rubin, Albert L.
: TITLE OF INVENTION: Methods Useful in Endotoxin
: TITLE OF INVENTION: Prophylaxis and Therapy
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Felfe & Lynch
: STREET: 805 Third Avenue
: CITY: New York City
: STATE: New York
```

US 09 240 146 2

Query Match 3.5% Score 7; DB 4; Length 729;

Best Local Similarity 100.0%; Pred. No. 62; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0;

QY 175 LAEKIKE 181

|||||

Db 448 LAEKIKE 454

RESULT 8

US-08 928 692 59

Sequence 59, Application US/09928692

Patent No. 5958727

GENERAL INFORMATION:

APPLICANT: Brody, Howard

APPLICANT: Yavet, Deborah S.

APPLICANT: Lamsa, Michael

APPLICANT: Hansen, Kim

TITLE OF INVENTION: Methods for Modifying the Production of

TITLE OF INVENTION: a Polypeptide

NUMBER OF SEQUENCES: 80

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 59587270 No. 5958727disk of No. 5958727th America, Inc.

STREET: 405 Lexington Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

FILING DATE: 12-SEPT-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4944,200-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 867-0123

TELEFAX: 212 867-9655

INFORMATION FOR SEQ ID NO: 59;

SEQUENCE CHARACTERISTICS:

LENGTH: 911 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 59587270

US-08-928-692-59

Query Match

Best Local Similarity 100.0%; Pred. No. 77; Length 911;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 LLLQFIL 80

|||||

Db 647 LLLQFIL 643

RESULT 9

US-08 928 692 58

Sequence 58, Application US/09928692

Patent No. 5958727

GENERAL INFORMATION:

APPLICANT: Brody, Howard

APPLICANT: Yavet, Deborah S.

APPLICANT: Lamsa, Michael,
 APPLICANT: Hansen, Kim
 TITLE OF INVENTION: Methods for Modifying the Production of
 TITLE OF INVENTION: a Polypeptide
 NUMBER OF SEQUENCES: 80
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 59587270 No. 5958727disk of No. 5958727th America, Inc.
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/928,692
 FILING DATE: 12-SEPT-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lambiris, Elias J.
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 4944,200-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-867-9655
 INFORMATION FOR SEQ ID NO: 58:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 916 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: No. 5958727e
 US-08-928-692-58

Query Match

Best Local Similarity 100.0%; Pred. No. 77; Length 916;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 LLLQFIL 80

|||||

Db 642 LLLQFIL 648

RESULT 10

US-07-596-467-6

Sequence 6, Application US/07596467

Patent No. 5306862

GENERAL INFORMATION:

APPLICANT: Chappell, J.

APPLICANT: Saunders, Court A.

APPLICANT: Cuellar, R.

APPLICANT: Wolf, Fred R.

TITLE OF INVENTION: Method and Composition for Increasing

TITLE OF INVENTION: Sterol Accumulation in Higher Plants

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dressler, Goldsmith, Shore, Sutker & Milnamow

STREET: 180 N. Stetson St.

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/596,467

CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSFO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/018,652
FILING DATE: 12-SEPT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4944,200-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 178 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5958727e
US-08-928-692-57

Query Match 3.5%; Score 7; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 LLLQFIL 80
Db 104 LLLQFIL 110

RESULT 5
US-09-347-833-6
; Sequence 6, Application US/09347833
; Patent No. 6294658
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; TITLE OF INVENTION: Factors Involved in Gene Expression
; FILE REFERENCE: BB-1172
; CURRENT APPLICATION NUMBER: US/09/347,833
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,415
; EARLIER FILING DATE: July 10, 1998
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Microsoft office 97
; SEQ ID NO 6
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Glycine max
US-09-347-833-6

Query Match 3.5%; Score 7; DB 4; Length 428;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 PLVDQLE 112
Db 357 PLVDQLE 363

RESULT 6
US-09-347-833-4
; Sequence 4, Application US/09347833

Patent No. 6294658
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Odell, Joan T.
; TITLE OF INVENTION: Factors Involved in Gene Expression
; FILE REFERENCE: BB-1172
; CURRENT APPLICATION NUMBER: US/09/347,833
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,415
; EARLIER FILING DATE: July 10, 1998
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Microsoft office 97
; SEQ ID NO 4
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-347-833-4

Query Match 3.5%; Score 7; DB 4; Length 655;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 PLVDQLE 112
Db 583 PLVDQLE 589

RESULT 7
US-09-230-196-2
; Sequence 2, Application US/09230196
; Patent No. 6307035
; GENERAL INFORMATION:
; APPLICANT: Kauscher III, Frank J.
; APPLICANT: Jensen, David E.
; TITLE OF INVENTION: BRCA1 Associated Protein (BAP-1) and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr., PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/230,196
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,997
; FILING DATE: 02-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/038,109
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WS168HUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 729 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

27 66 ENDEK 72

28 29 ENDEK 35

RESULT 2

US 08-211-202-140
Sequence 140, Application US/08211202

Patent No. 5565332

GENERAL INFORMATION:

APPLICANT: HODENBOM, Hendricus Renerus Jacobus Mattheus

APPLICANT: BAIER, Michael

APPLICANT: JESPERSEN, Laurent Stephane Anne Therese

APPLICANT: WINNER, Gregory Paul

TITLE OF INVENTION: Production of chimeric antibodies - a

TITLE OF INVENTION: combinatorial approach

NUMBER OF SEQUENCES: 144

CORRESPONDENCE ADDRESS:

ADDRESSEE: David W. Clough, Marshall Oppole Gerstein Murray &

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/211,202

FILING DATE: 23 SEP 1992

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: GB 9120253.3

FILING DATE: 23-SEP-1991

APPLICATION DATA:

APPLICATION NUMBER: GB 9120377.8

FILING DATE: 25-SEP-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9206318.9

FILING DATE: 24-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9206372.6

FILING DATE: 24-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FCT/GB92/00883

FILING DATE: 15-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: David W. Clough

REGISTRATION NUMBER: 36,107

REFERENCE NUMBER: 2811741960

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474 9300

TELEFAX: 312-474 0418

TELEX: 25-4856

INFORMATION FOR SEQ ID NO: 140:

SEQUENCE CHARACTERISTICS:

LENGTH: 98 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US 08-211-202-140

Query Match

Best Local Similarity 100.0%; Seq. ID No. 140; Length 98;

Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 120 KQSERVT 126

Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 63 KQSERVT 69

RESULT 3

US-08-545-809A-139

Sequence 139, Application US/08545809A

Patent No. 6096878

GENERAL INFORMATION:

APPLICANT: Honjo, Yasuko

APPLICANT: Matsuda, Fumihiko

TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE

NUMBER OF SEQUENCES: 145

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110 2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/545,809A

FILING DATE: 27-MAR-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/00603

FILING DATE: 10-MAY-1993

ATTORNEY/AGENT INFORMATION:

NAME: Freeman, John W.

REGISTRATION NUMBER: 29,066

REFERENCE/DOCKET NUMBER: 06501/004001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 139:

SEQUENCE CHARACTERISTICS:

LENGTH: 117 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-545-809A-139

Query Match 3.5%; Score 7: DB 4; Length 117;

Best Local Similarity 100.0%; Seq. ID No. 12; Length 117;

Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 120 KQSERVT 126

Db 82 KQSERVT 88

RESULT 4

US-08-928-692-57

Sequence 57, Application US/08928692

Patent No. 5958727

GENERAL INFORMATION:

APPLICANT: Brody, Howard

APPLICANT: Yaver, Deborah S.

APPLICANT: Lamsa, Michael

APPLICANT: Hansen, Kim

TITLE OF INVENTION: Methods for Modifying the Production of

NUMBER OF SEQUENCES: 80

CORRESPONDENCE ADDRESS:

ADDRESSEE: No 5958727, No. 5958727disk of No. 5958727th America, Inc.

STREET: 405 Lexington Avenue

QY 120 KQSERVT 126

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using SW model

Run on: July 2, 2002, 16:04:00 : Search time 44.85 seconds
(without alignments)
109,466 Million cell updates/sec

Title: US-09-603-665-5_COPY_1917_2117
Perfect score: 201
Sequence: 1 PKLFWAKTDAKQWLLTF NYIVLLPESLPPLAELEMEDE 201

Scoring table:
OLIGO
Gapop 60 0 : Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 6

Total number of hits satisfying chosen parameters: 176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pcp.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pcp.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pcp.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pcp.*
5: /cgn2_6/ptodata/2/1aa/PTLUS_COMB.pcp.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pcp.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

Result No.	Score	Query Match	Length	DB ID	Description
1	7	3.5	41	4	US-08-630-915A-79
2	7	3.5	98	1	US-08-211-202-130
3	7	3.5	117	3	US-08-545-809A-139
4	7	3.5	178	2	US-08-928-692-57
5	7	3.5	428	4	US-09-347-833-6
6	7	3.5	655	4	US-09-347-833-4
7	7	3.5	729	4	US-09-240-106-2
8	7	3.5	911	2	US-08-928-692-59
9	7	3.5	916	2	US-08-928-692-58
10	7	3.5	1045	1	US-07-596-467-6
11	7	3.5	1045	1	US-07-934-374-6
12	7	3.5	1045	1	US-07-783-861C-6
13	6	3.0	17	1	US-07-928-930A-3
14	6	3.0	17	1	US-08-288-568-3
15	6	3.0	17	1	US-08-487-461-3
16	6	3.0	17	1	US-08-432-691-3
17	6	3.0	17	1	US-08-487-459-3
18	6	3.0	18	1	US-07-928-930A-1
19	6	3.0	18	1	US-07-920-597-6
20	6	3.0	18	1	US-08-288-568-1
21	6	3.0	18	1	US-08-487-461-1
22	6	3.0	18	1	US-08-432-691-1
23	6	3.0	18	1	US-08-487-459-1
24	6	3.0	18	3	US-08-940-095-240
25	6	3.0	18	3	US-08-940-095-244
26	6	3.0	18	3	US-08-940-093-240
27	6	3.0	18	3	US-08-940-093-244

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/04/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-79
Query match 3.5%; Score 7; DB 4; length 41;
Best Local Similarity 100.0%; Pred. No. 4.4;

Query match 3.5% Score 7, DB 4, length 41;
Best Local Similarity 100.0%; Pred. No. 4.4;

ALIGNMENTS

RESULT 1
US-08-630-915A-79
; Sequence 79, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: POWELLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHOD OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-79

Sequence 240, App
Sequence 244, App
Sequence 240, App
Sequence 244, App
Sequence 240, App
Sequence 244, App
Sequence 240, App
Sequence 244, App
Sequence 22, Appl
Sequence 22, Appl
Sequence 12, Appl
Sequence 4, Appl
Sequence 110, App
Sequence 11, Appl
Sequence 7, Appl
Sequence 193, App
Sequence 193, App
Sequence 19, Appl
Sequence 19, Appl

```

RESULT 15
AAM69H27
ID AAM69H27 standard; protein; 110 AA.
XX
AC AAM69H27;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30133.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN W0200157276-A2.
XX
PD 09 AUG-2001.
XX
FE 10 JAN-2001; 2001WO-0500668.
XX
PR 04-FEB-2000; 2000US-0-80312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0508408.
PR 03-AUG-2000; 2000US-0632466.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000CH-0024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn St, Hanzel DK, Chen W, Rank DR;
XX
WP1: 2001-488400/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 30133; 658pp - Sequence listing: English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 110 AA:

```

```

Query Match 4.5%; Score 7; DB 22; Length 110;
Best local Similarity 100.0%; Prod. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

UY 173 LALAEL 179
bb 82 lalaekl 88

```

```

Search completed: July 2, 2002, 16:06:04
Job time: 181 sec

```

CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
 XX
 SQ Sequence 110 AA;

Query Match 3.5%, Score 7, DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 LALAEKL 179
 Db 82 lalaecl 88

RESULT 13
 ABB22005
 ID ABB22005 standard; Protein: 110 AA.
 AC ABB22005;
 DT 23-JAN-2002 (first entry)
 DE Protein #4004 encoded by probe for measuring heart cell gene expression.
 XX Human: gene expression; heart, microarray, vascular system;
 KW cardiovascular disease; hypertension, cardiac arrhythmia,
 KW congenital heart disease.
 XX Homo sapiens.
 OS
 XX
 XX
 PN WO200157274-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-080066.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 04-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0336459.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPT: 2001-488899/53.

Single exon nucleic acid probes for analyzing gene expression in human hearts -

Claim 15. SEQ ID NO 23775; 530pp; English.

The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21535-ABM41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 110 AA;

Query Match 3.5%, Score 7, DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 LALAEKL 179
 Db 82 lalaecl 88

RESULT 14
 AAM57432
 ID AAM57432 standard; Protein: 110 AA.

XX
 AC AAM57432;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 29537.
 DE Human: brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-0800667

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 04-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0336459.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPT: 2001-483446/52

XX Single exon nucleic acid probes for analyzing gene expression in human brains -
 XX Example 4; SEQ ID NO: 29537; 550pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.

XX Sequence 110 AA;

Query Match 3.5%, Score 7, DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 LALAEKL 179
 Db 82 lalaecl 88

PA (MED) MEDICAL RES COUNCIL.
 XX
 PI Baier M, Hoogenboom RM, Jespers LSAT, Winter GP;
 XX
 DE WPI: 1994 117544/14.
 XX
 PI Producing human antibody polypeptide dimer specific for antigen -
 PI comprises use of chain shuffling using phage expression, useful
 PT for reducing anti globulin responses in humans for increased
 PT human characteristics
 XX
 PS Example: Fig 11; 109pp; English.
 XX
 CC the sequence is that of the Dp-74 VHL gene heavy chain
 CC which may be used as part of a method of producing chimeric
 CC mouse-human antibodies or fragments which have the same
 CC binding specificity as a parent Ab but have increased human
 CC characteristics, preventing anti-globulin response in humans.
 XX
 SQ Sequence 89 AA:
 Query Match 3.5%; Score 7; DP 14; Length 89;
 Best Local Similarity 00.0%; Prod. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 120 KQEKVT 126
 DB 111111
 63 Ktqekvt 69
 RESULT 11
 ABB41456
 ID ABB41456 standard; Peptide: 110 AA.
 AC ABB41456;
 XX
 DE 01 FEB-2002 (first entry)
 XX
 DE Peptide #4107 encoded by breast cell single exon nucleic acid probe.
 XX
 KW Human; microarray; single exon probe; gene expression; breast,
 KW disease; cancer.
 XX
 OS Homo sapiens.
 XX
 FN W0200157271-A2.
 XX
 PD 09 AUG 2001.
 XX
 PF 30 JAN 2001; 2001WO-US00662.
 XX
 PR 04 FEB-2000; 2000US-0180312.
 PR 26 MAY-2000; 2000US-0207456.
 PR 30 JUN-2000; 2000US-0608408.
 PR 03 AUG-2000; 2000US-0642364.
 PR 21 SEP-2000; 2000US-0244687.
 PR 27 SEP-2000; 2000US-0246359.
 PR 04 OCT-2000; 2000GB-0024264.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DE WPI: 2001 496944/54.
 XX
 PI New spatially addressable set of single exon nucleic acid probes,
 PI useful for measuring gene expression in sample derived from human
 PI breast, comprises number of single exon nucleic acid probes -
 XX
 PS claim 27; SEQ ID NO 14424; 327pp; sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from Wipo at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 110 AA:
 Query Match 3.5%; Score 7; DP 22; Length 110;
 Best Local Similarity 100.0%; Prod. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 173 LALAEKL 179
 DB 82 lalaecl 88
 RESULT 12
 ABB36668
 ID ABB36668 standard; Peptide: 110 AA.
 AC ABB36668;
 XX
 DE 04 FEB-2002 (first entry)
 XX
 DE Peptide #4174 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 FN W0200157277-A2.
 XX
 PD 09 AUG-2001.
 XX
 PF 30 JAN-2001; 2001WO-US00669.
 XX
 PR 04 FEB-2000; 2000US-0180312.
 PR 26 MAY-2000; 2000US-0207456.
 PR 30 JUN-2000; 2000US-0608408.
 PR 03 AUG-2000; 2000US-0632466.
 PR 21 SEP-2000; 2000US-0244687.
 PR 27 SEP-2000; 2000US-0246359.
 PR 04 OCT-2000; 2000GB-0024264.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DE WPI: 2001-483447/52.
 XX
 PI Human genome-derived single exon nucleic acid probes useful for
 PI analyzing gene expression in human foetal liver -
 XX
 PS claim 27; SEQ ID NO 29303; 639pp; sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,

XX Peptide #8214 encoded by probe for measuring placental gene expression.
 DE Probe: microarray, human, placenta; antenatal diagnosis,
 KW genetic disorder.
 KW Homo sapiens.
 OS
 XX
 XX WO200157272-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US000663.
 DE
 XX 04-FEB-2000; 2000US-0180312.
 PR
 XX 26-MAY-2000; 2000US-0207456.
 PP
 XX 30-JUN-2000; 2000US-0600406.
 PR
 XX 03-AUG-2000; 2000US-0632366.
 PR
 XX 21-SEP-2000; 2000US-0234687.
 PR
 XX 27-SEP-2000; 2000US-0246359.
 PR
 XX 04-OCT-2000; 2000US-0024263.
 PA
 (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR,
 XX WPI: 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta.
 PT
 XX Claim 27; SEQ ID NO 34446; 654bp; English.
 XX The present invention relates to single exon nucleic acid probes (SENP;
 CC see AAI31315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX Sequence 36 AA:
 SQ
 Query Match 3.5%, Score 7; DB 22; Length 36;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 116 GGRKKFO 122
 DB 4 ggeckfq 10
 RESULT 9
 AAO01170
 ID AAO01170 standard; Protein; 83 AA.
 XX
 AC AAO01170;
 XX
 XX 06-NOV-2001 (first entry)
 DT
 XX Human polypeptide SEQ ID NO 15062.
 DE
 XX Human; cytokine, cell proliferation, cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KW tissue growth factor; immunomodulatory, cancer; leukaemia,
 KW nervous system disorders; arthritis; inflammation.
 XX Homo sapiens.
 OS
 XX WO200104845-A2.
 PN
 XX 07-SEP-2001.
 PD
 XX

PF 26-FEB-2001; 2001WO-US04927.
 XX
 XX 28-FEB-2000; 2000US-0515126.
 PR
 XX 18-MAY-2000; 2000US-0577409.
 PA
 (HYSB-) HYSRO INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 PJ
 XX WPI: 2001-514838/56.
 DR
 XX N-PSDB; AAI81101.
 DR
 XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders.
 PT
 XX Claim 20; SEQ ID NO 15062; 1399bp; Sequence Listing, English.
 PS
 XX The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA019410) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine like activities,
 CC e.g. stem cell growth factor activity, hematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 83 AA:
 Query Match 3.5%, Score 7; DB 22; Length 83;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 150 NYQILK 156
 DB 33 nyqilk 39
 RESULT 10
 AAR34271
 ID AAR34271 standard; Protein; 89 AA.
 XX
 AC AAR34271;
 XX
 XX 26-JUL-1993 (first entry)
 DT
 XX Dp-74 VHL gene heavy chain.
 DE
 XX Human; monoclonal antibody; anti globulin response;
 KW chimeric; mouse-human antibodies; antibody; prevention.
 KW Homo sapiens.
 OS
 XX WO9306213-A.
 PN
 XX 01-APP-1993.
 PT
 XX 23-SEP-1992; 92WO-GB01755
 PR
 XX 23-SEP-1991; 91GB-0020252.
 PR
 XX 25-SEP-1991; 91GB-0020377.
 PR
 XX 24-MAR-1992; 92GB-0006318.
 PR
 XX 24-MAR-1992; 92GB-0006372.
 PR
 XX 15-MAY-1992; 92WO-GB00883.
 XX
 XX (CAMB-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 PA

cc measuring and displaying gene expression in samples derived from human
cc fetal liver. The present sequence is a peptide encoded by a single exon
cc nucleic acid probe of the invention.
cc Note: The sequence data for this patent did not form part of the
cc printed specification, but was obtained in electronic format directly
cc from WIPO at http://wipo.int/pub/published_pct_sequences.

XX Sequence 36 AA:

Query Match 4.5% Score 7; DB 22; Length 36;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 GSEERKQ 122
Db 4 qgeekiq 10
|||||

RESULT 6
AAM61274
ID AAM61279 standard; Protein: 36 AA.
AC AAM61274;

DE 05 NOV 2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 33384.
XX Human: brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.

XX Homo sapiens

OS W0200157276-A2.

PN 09-AUG-2001.

PP 40 JAN 2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0 80312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632466.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024253.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-484446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains

XX Example 4: SEQ ID NO: 33384; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
cc probes which are derived from genomic sequences expressed in the human
cc brain. They can be used to measure gene expression in brain cell samples,
cc which may enable the diagnosis and improved treatment of nervous system
cc diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
cc epilepsy and cancers. The present sequence is a protein encoded by one of
cc the probes of the invention.

XX Sequence 36 AA:

Query Match 3.5% Score 7; DB 22; Length 36;
Best Local Similarity 100.0%; Pred. No. 9.7;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 116 GSEERKQ 122
Db 4 qgeekiq 10
|||||

RESULT 7
AAM74002
ID AAM74002 standard; Protein: 36 AA.

AC AAM74002;

DE 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 44408.
XX Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.

OS Homo sapiens.

PN W0200157276-A2.

PP 09-AUG-2001.

PR 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632466.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024253.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow.

XX Example 4: SEQ ID NO: 34303; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
cc probes which are derived from genomic sequences expressed in the human
cc bone marrow. They can be used to measure gene expression in bone marrow
cc samples, which may enable the improved diagnosis and treatment of cancers
cc such as lymphoma, leukaemia and myeloma. The present sequence is a
cc protein encoded by one of the probes of the invention.

XX Sequence 36 AA;

Query Match 3.5% Score 7; DB 22; Length 36;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 GSEERKQ 122
Db 4 qgeekiq 10
|||||

RESULT 8
AAM34177
ID AAM34177 standard; Protein: 36 AA.

AC AAM34177;

DE 17-OCT-2001 (first entry)

PF 19-SEP-1997; 97WO-US16842.
 XX
 PR 04-APR-1997; 97HS-0042985.
 PP 20-SEP-1996; 96HS-0025296.
 PP 04-APR-1997; 97HS-0042611.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Baer R, Bowcock AM;
 XX
 DR WPI: 1998-230317/20.
 DR N-PSDB; AAV24135.
 XX
 PT DNA sequence encoding HARD1, B123, BE2, BE14, BE31 or BE445 - which
 PT as breast cancer antigen, BRCAL, binding proteins are useful to
 PT identify patient having or at risk of developing cancer
 XX
 PS Disclosure, Page 287-288, 348pp; English.
 XX
 CC The sequence is that of a protein which can be used in the
 CC preparation of the recombinant breast cancer antigen, BRCAL, binding
 CC proteins HARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the
 CC detection of a HARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid
 CC sequence, specifically a wild type HARD1 composition for the detection
 CC or purification of BRCAL, useful to identify a patient having, or at
 CC risk of developing cancer. HARD1 can be used in the preparation of an
 CC anti-HARD1 antibody, and in the detection and purification of a BRCAL
 CC protein. HARD1, B123, BE2, BE14, BE31 or BE445 can be used in the
 CC identification of a binding protein agonist or antagonist that alters
 CC the binding of HARD1, B123, BE2, BE14, BE31 or BE445 to BRCAL or the
 CC biological activity of the BRCAL-HARD1, B123, BE2, BE14, BE31 or BE445
 CC complex. The antibodies can be used to detect HARD1, B123, BE2, BE14,
 CC BE31 or BE445, a specific anti-HARD1 antibody can be used to identify
 CC a patient having or at risk of developing cancer.
 XX
 SQ Sequence 515 AA;

 Query Match 33.3%; Score 67; DB 19; Length 515;
 Best Local Similarity 100.0%; Pred No. 1.7e-61;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 135 QFSVAMADSLWKPLNYQILIKTRUSSPKVRFPAALITVLALAEKIKENYIVLIPESIPPL 194
 Db 422 qtsvamaddslwkplnyqiliktusspkvrfpaalitvlalaeikenyivilpesipil 481
 QY 195 AELMEDE 201
 Db 482 acImede 488

 RESULT 4
 AAY23793
 ID AAY23793 standard: Protein, 220 AA.
 XX
 AC AAY23793;
 XX
 DT 14-SEP-1999 (first entry)
 XX
 DE A gyrase protein of Chitinophaga pinensis.
 XX
 KW Identification, detection, microbe, gyrase gene, gyrase protein.
 XX
 OS Chitinophaga pinensis.
 XX
 PN JP11169175-A.
 XX
 PD 29-JUN-1999.
 XX
 PF 12-DEC-1997; 97JP-0343316.
 XX
 PR 12-DEC-1997; 97JP-0343316.
 XX
 PA (MARL-) MARINE BIOTECHNOLOGY INST CO LTD.
 PA (KALY-) KALYO BIOTECHNOLOGY KENKYUSHO KK.
 WPI: 1999-422615/36.
 N-PSDB; AAX86000.
 identification and detection of a microbe - by detection of a
 gyrase gene
 Example 3; Page 14-15; 42pp; Japanese.
 The specification describes a method for the identification or
 detection of a microbe, using the gyrase gene as the index. The
 method involves the use of PCR primers to amplify DNA from the
 microbe, which is then identified or detected depending on its
 base sequence. The method can be used to classify and identify
 an unidentified microbe strain rapidly and with high precision.
 The present sequence represents a gyrase protein.
 Sequence 220 AA;

 Query Match 4.0%; Score 8; DB 20; Length 220;
 Best Local Similarity 100.0%; Pred No. 4.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 65 SENDPEKC 72
 Db 130 sendpekc 137

 RESULT 5
 ABB40457
 ID ABB40457 standard: Peptide; 36 AA.
 XX
 AC ABB40457;
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #7963 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression, single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 DT 09-AUG-2001.
 XX
 PD 30 JAN-2001; 2001WO-US00669
 XX
 DE 04-FEB-2000; 2000US-0180412
 PR 25-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000US-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzei HK, Chen W, Rank DR;
 XX
 DR WPI: 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human foetal liver -
 XX
 PS Claim 27; SEQ ID NO 33092; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,


```

XX Harry C. Bouquelet et al. Chumakov I., Cohen Akenine A;
XX WPI: 2001-467042/48.
XX N PSUB: AAF83409, AAF83410.
XX New HAP28 polynucleotides and polypeptides overexpressed in prostate
XX cancer cells for diagnosing prostate tumors, e.g. by hybridization or
XX polymerase chain reaction assays -
XX claim 14; Page 297-304; 349pp; English.
XX The invention is directed to HAP28 polypeptides, HAP28 polynucleotide
XX sequences and regulatory region located at the 3' and 5' ends of the
XX BAP28 coding region. The BAP28 polypeptides can be expressed by standard
XX recombinant methodology. BAP28 polynucleotides and polypeptides have been
XX found to be over expressed in prostate tumour cells, therefore levels of
XX HAP28 expression and/or activity may be assayed (e.g. by polymerase chain
XX reaction (PCR)) to diagnose patient suffering from or susceptible to
XX prostate cancer. Antibodies specific for the HAP28 polypeptides are
XX useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
XX useful in genetic analysis. The present sequence represents a protein
XX encoded by a first cDNA sequence of the BAP28 gene consisting of the
XX exons 1 to 45.
XX Sequence 2144 AA:
SQ
Query Match 99.0%, Score 155, DB 22, Length 2144.
Best Local Similarity 100.0%, Pred. No. 3,46 199.
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FKLFWAKTELAPKORLLTYNLDGIAEKKLGLFTLFLGHLVKKPFDLTXQVNSKTDE 60
DB 1617 LPLFLGAKKGLGKFLGLFLGHLVKKPFDLTXQVNSKTDE 1776
QY 61 AFDSNMPKQVLLQFLLNLYKIFLFDGHPISKERAXALMPLVQLDMRLGGEKK 120
DB 1977 ALDSDNDPEKCELLQLILNLYKILFLDGLHLSKERAXALMPLVQLDMRLGGEKK 2036
QY 121 FQFVTKHLIPICIAFVSAMADSLWKPFLNYQIITWTPSSPVPFPAALITVLALAEK 180
DB 2037 FQFVTKHLIPICIAFVSAMADSLWKPFLNYQIITWTPSSPVPFPAALITVLALAEK 2096
QY 181 ENYVILPESIFLAELMEDE 201
DB 2097 ENYVILPESIFLAELMEDE 2117
RESULTS 2
AAF92729
ID AAF92729 standard; Protein: 349 AA.
XX AAF92729
XX AAF92729:
XX 26 JUN 2001 (first entry)
XX Human protein sequence SEQ ID NO:11159.
XX Human; primer: detection; diagnosis: antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617 A2.
XX 07 FEB 2001.
XX 28 JUL 2000; 2000JP-0116126.
XX 29 JUL 1999; 99JP-0248036.
XX 27 AUG 1999; 99JP-0300254.
XX 11 JAN 2000; 2000JP-0118776.
XX 02 MAY 2000; 2000JP-0184767.

```

```

PR 09-JUN-2000; 2000JP-0241899.
XX (HELL-) HELIX RES INST.
XX Ota T., Isozaki T., Nishikawa T., Hayashi K., Saito K., Yamamoto J.;
XX Ishii S., Sugiyama T., Wakamatsu A., Nagai K., Otsuki I.;
XX WPI: 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX claim 8; SEQ ID 11159; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification, the primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAF03166 to AAF13628 and
XX AAF13633 to AAF18742 represent human cDNA sequences, AAF92446 to
XX AAF95893 represent human amino acid sequences; and AAF13629 to AAF13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX Sequence 349 AA:
SQ
Query Match 49.8%, Score 190, DB 22, Length 349;
Best Local Similarity 100.0%, Pred. No. 3,16 96;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 102 ALMPLVDLENLRLGGEKFEQVTKHLIPICIAFVSAMADSLWKPFLNYQIITWTPSS 161
DB 223 almpplvdqlenrlggekfeqvtkhlipiciafvsamaddslwkpflnyqiitwtrdss 282
QY 162 PKYRFAALITVLALAEKLENYVILPESIFLAELMEDE 201
DB 283 pkvriaalitvialaeklenvivilpesiflaelmede 322
RESULTS 3
AAF54099
ID AAF54099 standard; Protein 515 AA.
XX AAF54099;
XX 28-Sep-1998 (first entry)
XX Homo sapiens BAP28 sequence.
XX DE HAR01; ring protein; BRCA1, breast cancer; risk; diagnosis.
XX KW Homo sapiens.
XX OS
XX W00912327-A2.
XX 26-MAR-1998.
XX

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw mode.)

Run on: July 2, 2002, 16:06:03 ; Search time 117.59 Seconds
(without alignments)
189,862 Million cell updates/sec

Title: US-09-603-665-5_copy_1917_2117
Perfect score: 201
Sequence: 1 PKLFWAKTEAPKDRILTF.....NYIVLPSIPPLAFIMDE 201

Scoring table: Oligo
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 6

Total number of hits satisfying chosen parameters: 729

Minimum DB seq length: 6

Maximum DB seq length: 2000000000

Post-processing. Listing first 45 summaries

Database : A_Genescq_032802.*

- 1: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1980.DAT.*
- 2: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1981.DAT.*
- 3: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1982.DAT.*
- 4: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1983.DAT.*
- 5: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1984.DAT.*
- 6: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1985.DAT.*
- 7: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1986.DAT.*
- 8: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1987.DAT.*
- 9: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1988.DAT.*
- 10: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1989.DAT.*
- 11: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1990.DAT.*
- 12: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1991.DAT.*
- 13: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1992.DAT.*
- 14: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1993.DAT.*
- 15: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1994.DAT.*
- 16: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1995.DAT.*
- 17: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1996.DAT.*
- 18: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1997.DAT.*
- 19: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1998.DAT.*
- 20: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA1999.DAT.*
- 21: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA2000.DAT.*
- 22: /SIDSI/qcdata/hold-geneseq/genescq-emb1/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	199	99.0	2144	22	AAH85029
2	103	49.8	349	22	AAH92719
3	67	33.3	515	19	AAW54099
4	4	4.0	220	20	AAH23743
5	7	3.5	36	22	AAH40437
6	7	3.5	36	22	AAW61274
7	7	3.5	36	22	AAW74002
8	7	3.5	36	22	AAW34177
9	7	3.5	83	22	AAW01170
10	7	3.5	89	14	AAK34271
11	7	3.5	110	22	AAH31456

12	7	3.5	110	22	ABH36668
13	7	3.5	110	22	ABH22005
14	7	3.5	110	22	AAW57432
15	7	3.5	110	22	AAW69827
16	7	3.5	110	22	AAW17651
17	7	3.5	110	22	AAW30169
18	7	3.5	110	22	AAW05309
19	7	3.5	117	16	AAW66345
20	7	3.5	165	21	AAH27313
21	7	3.5	178	19	AAW37995
22	7	3.5	178	20	AAW49876
23	7	3.5	200	21	AAW34112
24	7	3.5	276	18	AAW55542
25	7	3.5	277	18	AAW55442
26	7	3.5	342	21	AAW42697
27	7	3.5	342	22	AAW66797
28	7	3.5	353	21	AAW43736
29	7	3.5	353	22	AAW73794
30	7	3.5	367	21	AAW66633
31	7	3.5	367	22	AAW65156
32	7	3.5	374	22	AAW65309
33	7	3.5	395	19	AAW57571
34	7	3.5	437	20	AAW37663
35	7	3.5	521	22	AAW59537
36	7	3.5	604	21	AAW27324
37	7	3.5	655	22	AAW13001
38	7	3.5	729	19	AAW47538
39	7	3.5	1045	15	AAW58611
40	7	3.5	1122	16	AAW64037
41	7	3.5	1138	21	AAW81512
42	7	3.5	1216	22	AAW37719
43	7	3.5	1216	22	AAW38105
44	6	3.0	9	18	AAW31645
45	6	3.0	12	18	AAW31644

ALIGNMENTS

RESULT 1	
AAH85029	
ID	AAH85029 standard; protein; 2144 AA.
XX	AAH85029;
XX	06-AUG-2001 (first entry)
DE	Protein encoded by BAP28 cDNA consisting of exons 1 to 45.
DE	BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.
KW	Homo sapiens.
OS	Homo sapiens.
XX	
PH	Key location/Qualifiers
FT	Misc-difference 1694
FT	/label= Ser or Asn
FT	Misc-difference 1854
FT	/label= Ala or Val
FT	Misc-difference 1967
FT	/label= Asp or Asn
FT	Misc-difference 2017
FT	/label= Gly or Glu
XX	W0200100669-A2
PN	04-JAN-2001.
XX	
XX	23 JUN 2000. 2000000-1001183
XX	25-JUN-1999. 90005-0141323.
PR	18-JAN-2000; 2000005-0176880.
XX	(GENSET) GENSET.
PA	

1967-56
2017-96

Query Match 4.5%, Score 7; DB 2; Length 56;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 131 NQLEEVG 137
 L L L L L L
 Db 38 NQLEEVG 44

Search completed: July 2, 2002, 16:18:30
 Job time: 862 sec

DR Pfam: PF01585, G patch; 1.
 DR Pfam: PF01805; Surp; 1.
 DR SMART: SM00443; G-patch; 1.
 SQ SEQUENCE 884 AA; 100015 MW; 518HFP371BA132D9 CRC64;

Query Match 4.0%; Score 8; DB 4; Length 884;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LAALQKW 15
 |||||
 DB 202 LAALQKW 209

RESULT 13

ID Q21325 PRELIMINARY; PRT; 895 AA

AC Q21325; Q94264.

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE HYPOTHETICAL PROTEIN K08D10.1.

GN K08D10.1

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhaditida; Rhaditidae.

OC Rhaditidae; Plectoderinae; Caenorhabditis.

OX NCBI_TaxID 6239.

RN |||

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

PX MEDLINE=99069613; PubMed=9841916;

RA None;

RT "Genome sequence of the nematode *C. elegans*: a platform for

investigating biology, the *C. elegans* Sequencing Consortium.";

PL Science 282:2012-2018 (1998)

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Geisel C.; Bradshaw H.;

RT "The sequence of *C. elegans* cosmid K08D10.1."

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RT "Direct Submission."

PL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: H55857; AAA98035.2; -.

DR EMBL: U70855; AAB09160.2; -.

SQ SEQUENCE 895 AA; 103130 MW; E5APD7041235037 CRC64;

Query Match 4.0%; Score 8; DB 5; Length 895;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 ENDLEEVG 137
 |||||
 DB 696 ENDLEEVG 703

RESULT 14

Q9EWA1

ID Q9EWA1 PRELIMINARY; PRT; 9507 AA.

AC Q9EWA1.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE PIMS2 PROTEIN.

GN PIMS2.

OS Streptomyces natalensis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID: 68242;

RN |||

RP SEQUENCE FROM N.A.

RC MEDLINE=20547809; PubMed=11094342;

RA Aparicio J.F.; Fouces R.; Mendes M.V.; Olivera N.; Martin J.F.;

RT "A complex multienzyme system encoded by five polyketide synthase

genes is involved in the biosynthesis of the 26-membered polyene

macrolide pimaricin in *Streptomyces natalensis*."

RJ Chem Biol 7:495-505 (2000).

DR EMBL: AJ278573; CAC20921.1; -.

DR HSSP: P25715; IMLA.

DR InterPro: IPR001227; Acyltransf_domain.

DR InterPro: IPR000794; Ketoacyl-synt.

DR InterPro: IPR003880; Phosphonat_attach.

DR Pfam: PF00698; Acyl_transf; 6.

DR Pfam: PF00109; Ketoacyl-synt; 6.

DR Pfam: PF02801; ketoacyl-synt_C; 6

DR Pfam: PF00550; pp-binding; 6

DR PROSITE: PS00075; ACP_DOMAIN; 6.

DR PROSITE: PS00012; PHOSPHONATETHEINE; 1.

DR Phosphonatetheine.

SW SEQUENCE 9507 AA; 994228 MW; 57HD80C8AD37F89F CRC64;

Query Match 4.0%; Score 8; DB 2; Length 9507;
 Best Local Similarity 100.0%; Pred. No. 2; 4e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAAL 11
 |||||

DB 7587 LLSALAAL 7594

RESULT 15

Q9X617

ID Q9X617 PRELIMINARY; PRT; 56 AA.

AC Q9X617; 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE ALANYL-TRNA SYNTHETASE (FRAGMENT).

GN ALAS.

OS Flavobacterium ferrugineum.

OC Bacteria; CF8 group; Flavobacteria; Flavobacteriaceae; Flavobacterium.

OX NCBI_TaxID=249;

RN |||

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC13524;

PX MEDLINE=99058143; PubMed=9841678;

RA Gupta R.S.;

RT "Protein phylogenies and signature sequences: A reappraisal of

evolutionary relationships among archaebacteria, eubacteria, and

eukaryotes."

RT Microbiol Mol Biol Rev 62:1435-1491 (1998).

PL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC13524;

RA Gupta R.S.; Mukhtar T.; Singh B.;

RT "Evolutionary relationships among photosynthetic prokaryotes

(Helio bacterium chlorum, Chloroflexus auranticus, Cyanobacteria,

Chlorobium tepidum and Proteobacteria): implications regarding the

origin of photosynthesis."

RT Mol. Microbiol. 0:0-0 (1999).

DR EMBL: AF130451; AA033469.1; -.

DR InterPro: IPR002318; tPNA-synt_2C.

DR Pfam: PF01411; tRNA-synt_2c; 1.

DR Aminoacyl-tRNA synthetase.

FT NON_TER 1

FT NON_TER 56

FT SEQUENCE 56 AA; 6451 MW; 8623CEP781D1B4E7 CRC64;

RT The clobA genes from pseudomonas aeruginosa code for a novel cyanide-insensitive terminal oxidase related to the cytochrome bd quinol oxidases.*

Mol. Microbiol. 24:574-591(1997).

121

RP SEQUENCE FROM N.A.

RC STRAIN ATCC 15692 / PA-01

RE MEDLINE-20447337; PubMed-10984043;

KA Stover C.K., Pham X.-Q. T., Erwin A.L., Hoffnagle W.O., Kowalik D.J., Lagrou M.,

KA Hickey M.J., Brinkman P.S.L., Hoffnagle W.O., Kowalik D.J., Lagrou M.,

KA Garber R.L., Collier A., Robinson P., Westbrock-Wadman S., Yuan Y.,

KA Brody L.L., Coulter S.N., Folsger F.P., Kas A., Larbig K., Lim P.M.,

KA Smith K.A., Spencer D.H., Wang G.F.-S., Wu Z., Paulsen T.T.,

KA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V., an

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.*

RL Nature 406:959-964(2000).

DR EMBL: Y10528; CAA71559.1;

DR EMBL: AE004810; AAC07417.1;

DR InterPro: IPR002585; Bae_084_Cox.

DR Pfam: PF01654; Bae_084_Cox; 1.

KW Complete proteome.

SQ SEQUENCE 488 AA; 5/235 MW; 40A342106B767553 CRC64;

Query Match 4.0% Score 8; DB 16; Length 488;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 LTAFLEA 121

11111111

DB 103 LTAFLEA 110

RESULT 10

Q92S67

ID Q92S67

AC Q92S67

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE PUTATIVE QUINOL OXIDASE SUBUNIT I TRANSMEMBRANE PROTEIN.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OX Rhizobiaceae; Sinorhizobium.

NCBI_TaxID=382;

SEQUENCE FROM N.A.

STRAIN-1021;

MEDLINE-21368234; PubMed-11474104;

KA Gallibert F., Finan T.M., Hong S.R., Puchler A., Abola P., Anpe F.,

KA Ratliff-Hubler F., Barnett M.J., Becker A., Hovstad P., Hoth G.,

KA Houtry M., Bowser L., Huhrmester J., Gadieu E., Capela D., Chain P.,

KA Gowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,

KA Gloux S., Gledie T., Giffen A., Golding B., Gouzy J., Gurjal M.,

KA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,

KA Kahn M.L., Kalmann S., Keating J.C.H., Kiss K., Komp C., Leclaire V.,

KA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,

KA Ransporger U., Surzycki F., Thebaud P., Vandenberg M.,

KA Vorholster F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.,

RT "The composite genome of the legume symbiont Sinorhizobium meliloti".

RL Science 293:668-672(2001).

DR EMBL: AL591784; CAC45131.1;

KW Complete proteome.

SQ SEQUENCE 493 AA; 54/281 MW; F9FA0F439H52AC31 CRC64;

Query Match

Best Local Similarity 1.0%; Score 8; DB 16; Length 493;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 LTAFLEA 121

11111111

Db 127 LTAFLEA 134

RESULT 11

Q95XA4

ID Q95XA4

AC Q95XA4

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE HYPERHETICAL 82.8 KDa PROTEIN.

OS Y20F4.5.

OC Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.

OC Rhabditidae; Telodoridae; Caenorhabditis.

NCBI_TaxID=6239;

SEQUENCE FROM N.A.

STRAIN-BRISTOL N2;

MEDLINE-99069613; PubMed-9351916;

None;

"Genome sequence of the nematode C. elegans: a platform for

RT investigating biology. The C. elegans Sequencing Consortium.*

RL Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

STRAIN-BRISTOL N2;

WA Waterston R.J.

"Direct Submission".

PL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AAC93703; AAL00865.1;

KW Hypothetical protein.

SQ SEQUENCE 725 AA; 82788 MW; 18DC466B4C9D0A35 CRC64;

Query Match

Best Local Similarity 4.0% Score 8; DB 5; Length 725;

Matches 8; Conservative 100.0%; Pred. No. 27;

Mismatches 0; Indels 0; Gaps 0;

QY 130 ENDEEVG 137

11111111

DB 523 ENDEEVG 530

RESULT 12

Q00302

ID Q00302

AC Q00302

DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE EPPROT 213-21.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

SEQUENCE FROM N.A.

MEDLINE-20256781; PubMed-10794731;

KA Laplanté J.M., O'Rourke F., Lu X., Fein A., Olsen A., Feinstein M.B.,

RT "Cloning of human Ca2+ homeostasis endoplasmic reticulum protein

RT (CHERP): regulated expression of antisense cDNA depletes CHERP,

RT inhibits intracellular Ca2+ mobilization and decreases cell

RT proliferation.*

RL Biochem. J. 348:189-194(2000).

SEQUENCE FROM N.A.

KA Laplanté J., O'Rourke F.A., Lu X., Matthews E., Olsen A., Choi J.S.,

KA Rose E., Feinstein M.B.

Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: U94836; AAB53327.1;

DR InterPro: IPR000467; G_patch.

DR InterPro: IPR000061; Surp.

SEQUENCE 343 AA; 36355 MW; F18ARCEAD31AD499 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 343;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LTAFLEA 11

Db 255 LTAFLEA 262

RESULT 6

ID Q98DW1 PRELIMINARY; PRT; 453 AA.

AC Q98DW1

DT 01-OCT-2001 (TREMREL 18, Created)

DT 01-OCT-2001 (TREMREL 18, Last sequence update)

DT 01-OCT-2001 (TREMREL 18, Last annotation update)

DE QUINOL OXIDASE SUBUNIT 1;

GN MLR4524.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.

OC Phyllobacteriaceae; Mesorhizobium.

OX NCHI_TaxID=381;

RN 111

RP SEQUENCE FROM N.A.

RC STRAIN-WAFF303099;

RX MEDLINE-21042430; PubMed 11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpou S., Sugimoto M.,

RA Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti.";

RL Insa Res. 7:331-338(2000).

DR EMBL; AP003004; HAF51159.1;

DR InterPro; IPR002585; Bac_Obq_Cox.

DR Pfam; PF01654; Bac_Obq_Cox; 1.

KW Complete Proteome.

SQ SEQUENCE 453 AA; 49427 MW; 1D2439B6E34HFAUF CRC64;

Query Match 4.0%; Score 8; DB 16; Length 453;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 114 LTAFLEA 121

Db 100 LTAFLEA 107

RESULT 7

ID Q86080

AC Q86080 PRELIMINARY; PRT; 465 AA.

DT 01-NOV-1998 (TREMREL 08, Created)

DT 01-NOV-1998 (TREMREL 08, Last sequence update)

DT 01-DEC-2001 (TREMREL 19, Last annotation update)

DE QUINOL OXIDASE SUBUNIT 1 OXA.

GN OXA.

OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).

OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;

OC Rhodobacter.

OX NCHI_TaxID=1063;

RN 111

RP SEQUENCE FROM N.A.

RX STRAIN-2.4.1T;

RX MEDLINE-20519422; PubMed=11064196;

RA Mouncey N.J., Gak E., Choudhary M., Oh J., Kaplan S.;

RT "Respiratory pathways of Rhodobacter sphaeroides 2.4.1(T)."

RT identification and characterization of genes encoding quinol

RT Oxidases.";

RL FEMS Microbiol. Lett. 192:205-210(2000).

DR EMBL; AF084032; AAC34673.1;

DR InterPro; IPR002585; Bac_Obq_Cox.

DR Pfam; PF01654; Bac_Obq_Cox; 1.

SQ SEQUENCE 465 AA; 51345 MW; 5C9F8091F5670410 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 465;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 114 LTAFLEA 121

Db 101 LTAFLEA 108

RESULT 8

ID Q99492

AC Q99492 PRELIMINARY; PRT; 469 AA.

DT 01-MAY-1997 (TREMREL 03, Created)

DT 01-MAY-1997 (TREMREL 03, Last sequence update)

DT 01-DEC-2001 (TREMREL 19, Last annotation update)

DE DAN26 PROTEIN (FRAGMENT).

GN DAN26.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCHI_TaxID=9606;

RN 111

RP SHORINCE FROM N.A.

RX MEDLINE-97051922; PubMed=8896557;

RA Imbert G., Saudou F., Yvert G., Devys D., Trottier Y., Garnier J.M.,

RA Weber C., Mandel J.L., Cancell G., Abbas N., Didierjean O.,

RA Stevanin G., Agid Y., Brice A.;

RT "Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with

RT high sensitivity to expanded CAG/glutamine repeats.";

RL Nat. Genet. 14:285-291(1996).

DR EMBL; Y08264; CAA84591.1;

DR InterPro; IPR000061; Surp.

DR Pfam; PF01805; Surp; 1.

FT NON_TER 1

SQ SEQUENCE 469 AA; 52531 MW; 690E2987AFD58FE CRC64;

Query Match 4.0%; Score 8; DB 4; Length 469;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LALQKV 15

Db 244 LALQKV 251

RESULT 9

ID Q07440

AC Q07440 PRELIMINARY; PRT; 488 AA.

DT 01-JUL-1997 (TREMREL 04, Created)

DT 01-JUL-1997 (TREMREL 04, Last sequence update)

DT 01-DEC-2001 (TREMREL 19, Last annotation update)

DE CYANIDE INSENSITIVE TERMINAL OXIDASE.

GN CIAA OR PA3930.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCHI_TaxID=287;

RN 111

RP SEQUENCE FROM N.A.

RX STRAIN-ATCC 15692 / PA01;

RX MPI INF-97323403; PubMed=9179851;

RA Cunningham L., Pitt M., Williams H.D.;

01 MAY-2000 (TRMBHrel. 13, Last sequence update)
 01-MAR-2001 (TRMBHrel. 19, Last annotation update)
 DE HYPOPHYSICAL 12.7 KDA PROTEIN.
 GN DR0207.
 OS Deinoceriscus radiodurans.
 CC Bacteria; Thermus/Saibacoccus group; Deinococcus, Deinococcus us.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-R1;
 RX MEDLINE-25046896; PubMed 10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Hair D.H., Winn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Liang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Gitterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salter S., Smith H.G., Venter A.,
 RA Fraser C.M.;
 RA "Genome sequence of the radioresistant bacterium Deinococcus
 RF radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001483; AAF09801.1;
 DR TIGR: DR0207;
 KW Hypothetical protein; complete proteome.
 SQ SEQUENCE 123 AA: 12733 MW: 38938730R30R3081C CRC64;

Query Match 4.0% Score 8; DB 15; Length 123;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAAL 11
 1111111
 DB 5 LLSALAAL 12

RESULT 3
 Q950F5 PRELIMINARY: PRI: 300 AA.
 AC Q950F5;
 DT 01-DEC-2001 (TRMBHrel. 19, Created)
 DT 01-DEC-2001 (TRMBHrel. 19, Last sequence update)
 DT 01-DEC-2001 (TRMBHrel. 19, Last annotation update)
 DE CYTOCHROME B (FRAGMENT).
 OS Rana luteiventris.
 CC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.
 OX NCBI_TaxID=58174;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21306528; PubMed-11412371;
 RA Bos D.H., Siles J.W., Jr.;
 RA "Phylogenetic and conservation genetics of the Columbia Spotted Frog
 (Rana luteiventris)."
 RL Mol. Ecol. 10:1499-1513(2001).
 DR EMBL: AY016673; AAK54420.1;
 KW Mitochondrion.
 FT NON_TER 1 1
 FT NON_TER 300 300
 SQ SEQUENCE 300 AA: 33374 MW: 212A26A6CA86BBB7 CRC64;

Query Match 4.0% Score 8; DB 8; Length 300;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAAL 11
 1111111
 DB 210 LLSALAAL 217

RESULT 4

Q94PH4 PRELIMINARY: PRI: 300 AA.
 AC Q94PH4;
 DT 01-DEC-2001 (TRMBHrel. 19, Created)
 DT 01-DEC-2001 (TRMBHrel. 19, Last sequence update)
 DT 01-DEC-2001 (TRMBHrel. 19, Last annotation update)
 DE CYTOCHROME B (FRAGMENT).
 OS Rana luteiventris.
 CC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.
 OX NCBI_TaxID=58174;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21306528; PubMed-11412371;
 RA Bos D.H., Siles J.W., Jr.;
 RA "Phylogenetic and conservation genetics of the Columbia Spotted Frog
 (Rana luteiventris)."
 RL Mol. Ecol. 10:1499-1513(2001).
 DR EMBL: AY016649; AAK54396.1;
 DR EMBL: AY016652; AAK54399.1;
 DR EMBL: AY016654; AAK54401.1;
 KW Mitochondrion.
 FT NON_TER 1 1
 FT NON_TER 300 300
 SQ SEQUENCE 300 AA: 33606 MW: 562A26BB7DBF5CA3 CRC64;

Query Match 4.0% Score 8; DB 8; Length 300;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAAL 11
 1111111
 DB 210 LLSALAAL 217

RESULT 5
 Q9K3Y1 PRELIMINARY: PRI: 343 AA.
 AC Q9K3Y1;
 DT 01-OCT-2000 (TRMBHrel. 15, Created)
 DT 01-OCT-2000 (TRMBHrel. 15, Last sequence update)
 DT 01-JUN-2001 (TRMBHrel. 17, Last annotation update)
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR.
 GN ZSC061.35.
 OS Streptomyces coelicolor.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN A3(2);
 RA Oliver K., Harris D.;
 RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D., Parkhill J., Barrell R.G., Rajandream M.A.;
 RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE-97000551; PubMed-8843436;
 RA Pedenbach M., Kleser H.M., Denapalle D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RA "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL359949; CAB95810.1;
 DR InterPro: IPR002197; HTH_Fis.
 DR InterPro: IPR000792; HTH_boxR.
 DR SMART: SM00421; HTH_JUXX; 1.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model
Run on: July 2, 2002, 16:18:28 ; Search time 96.19 Seconds
(without alignments)
361.493 Million cell updates/sec

Title: US-09-603-665-5_COPY_1754_1954
Perfect score: 201
Sequence: 1 EVYLLSAIAAIQKVVETLPH.....TFYNIADCIKAKLKGILTLF 201

Scoring table: OLI60
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 17294229 residues

Word size : 6

Total number of hits satisfying chosen parameters: 2002

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:
1: sp-archaea:
2: sp-bacteria:
3: sp-fungi:
4: sp-human:
5: sp-invertebrate:
6: sp-mammal:
7: sp-mhc:
8: sp-organism:
9: sp-phage:
10: sp-plant:
11: sp-rodent:
12: sp-virus:
13: sp-vertebrate:
14: sp-unclassified:
15: sp-virus:
16: sp-bacteriophage:
17: sp-archaea:
18: sp-bacteria:
19: sp-fungi:
20: sp-human:
21: sp-invertebrate:
22: sp-mammal:
23: sp-mhc:
24: sp-organism:
25: sp-phage:
26: sp-plant:
27: sp-rodent:
28: sp-virus:
29: sp-vertebrate:
30: sp-unclassified:
31: sp-virus:
32: sp-bacteriophage:
33: sp-archaea:
34: sp-bacteria:
35: sp-fungi:
36: sp-human:
37: sp-invertebrate:
38: sp-mammal:
39: sp-mhc:
40: sp-organism:
41: sp-phage:
42: sp-plant:
43: sp-rodent:
44: sp-virus:
45: sp-vertebrate:
46: sp-unclassified:
47: sp-virus:
48: sp-bacteriophage:
49: sp-archaea:
50: sp-bacteria:
51: sp-fungi:
52: sp-human:
53: sp-invertebrate:
54: sp-mammal:
55: sp-mhc:
56: sp-organism:
57: sp-phage:
58: sp-plant:
59: sp-rodent:
60: sp-virus:
61: sp-vertebrate:
62: sp-unclassified:
63: sp-virus:
64: sp-bacteriophage:
65: sp-archaea:
66: sp-bacteria:
67: sp-fungi:
68: sp-human:
69: sp-invertebrate:
70: sp-mammal:
71: sp-mhc:
72: sp-organism:
73: sp-phage:
74: sp-plant:
75: sp-rodent:
76: sp-virus:
77: sp-vertebrate:
78: sp-unclassified:
79: sp-virus:
80: sp-bacteriophage:
81: sp-archaea:
82: sp-bacteria:
83: sp-fungi:
84: sp-human:
85: sp-invertebrate:
86: sp-mammal:
87: sp-mhc:
88: sp-organism:
89: sp-phage:
90: sp-plant:
91: sp-rodent:
92: sp-virus:
93: sp-vertebrate:
94: sp-unclassified:
95: sp-virus:
96: sp-bacteriophage:
97: sp-archaea:
98: sp-bacteria:
99: sp-fungi:
100: sp-human:
101: sp-invertebrate:
102: sp-mammal:
103: sp-mhc:
104: sp-organism:
105: sp-phage:
106: sp-plant:
107: sp-rodent:
108: sp-virus:
109: sp-vertebrate:
110: sp-unclassified:
111: sp-virus:
112: sp-bacteriophage:
113: sp-archaea:
114: sp-bacteria:
115: sp-fungi:
116: sp-human:
117: sp-invertebrate:
118: sp-mammal:
119: sp-mhc:
120: sp-organism:
121: sp-phage:
122: sp-plant:
123: sp-rodent:
124: sp-virus:
125: sp-vertebrate:
126: sp-unclassified:
127: sp-virus:
128: sp-bacteriophage:
129: sp-archaea:
130: sp-bacteria:
131: sp-fungi:
132: sp-human:
133: sp-invertebrate:
134: sp-mammal:
135: sp-mhc:
136: sp-organism:
137: sp-phage:
138: sp-plant:
139: sp-rodent:
140: sp-virus:
141: sp-vertebrate:
142: sp-unclassified:
143: sp-virus:
144: sp-bacteriophage:
145: sp-archaea:
146: sp-bacteria:
147: sp-fungi:
148: sp-human:
149: sp-invertebrate:
150: sp-mammal:
151: sp-mhc:
152: sp-organism:
153: sp-phage:
154: sp-plant:
155: sp-rodent:
156: sp-virus:
157: sp-vertebrate:
158: sp-unclassified:
159: sp-virus:
160: sp-bacteriophage:
161: sp-archaea:
162: sp-bacteria:
163: sp-fungi:
164: sp-human:
165: sp-invertebrate:
166: sp-mammal:
167: sp-mhc:
168: sp-organism:
169: sp-phage:
170: sp-plant:
171: sp-rodent:
172: sp-virus:
173: sp-vertebrate:
174: sp-unclassified:
175: sp-virus:
176: sp-bacteriophage:
177: sp-archaea:
178: sp-bacteria:
179: sp-fungi:
180: sp-human:
181: sp-invertebrate:
182: sp-mammal:
183: sp-mhc:
184: sp-organism:
185: sp-phage:
186: sp-plant:
187: sp-rodent:
188: sp-virus:
189: sp-vertebrate:
190: sp-unclassified:
191: sp-virus:
192: sp-bacteriophage:
193: sp-archaea:
194: sp-bacteria:
195: sp-fungi:
196: sp-human:
197: sp-invertebrate:
198: sp-mammal:
199: sp-mhc:
200: sp-organism:
201: sp-phage:
202: sp-plant:
203: sp-rodent:
204: sp-virus:
205: sp-vertebrate:
206: sp-unclassified:
207: sp-virus:
208: sp-bacteriophage:
209: sp-archaea:
210: sp-bacteria:
211: sp-fungi:
212: sp-human:
213: sp-invertebrate:
214: sp-mammal:
215: sp-mhc:
216: sp-organism:
217: sp-phage:
218: sp-plant:
219: sp-rodent:
220: sp-virus:
221: sp-vertebrate:
222: sp-unclassified:
223: sp-virus:
224: sp-bacteriophage:
225: sp-archaea:
226: sp-bacteria:
227: sp-fungi:
228: sp-human:
229: sp-invertebrate:
230: sp-mammal:
231: sp-mhc:
232: sp-organism:
233: sp-phage:
234: sp-plant:
235: sp-rodent:
236: sp-virus:
237: sp-vertebrate:
238: sp-unclassified:
239: sp-virus:
240: sp-bacteriophage:
241: sp-archaea:
242: sp-bacteria:
243: sp-fungi:
244: sp-human:
245: sp-invertebrate:
246: sp-mammal:
247: sp-mhc:
248: sp-organism:
249: sp-phage:
250: sp-plant:
251: sp-rodent:
252: sp-virus:
253: sp-vertebrate:
254: sp-unclassified:
255: sp-virus:
256: sp-bacteriophage:
257: sp-archaea:
258: sp-bacteria:
259: sp-fungi:
260: sp-human:
261: sp-invertebrate:
262: sp-mammal:
263: sp-mhc:
264: sp-organism:
265: sp-phage:
266: sp-plant:
267: sp-rodent:
268: sp-virus:
269: sp-vertebrate:
270: sp-unclassified:
271: sp-virus:
272: sp-bacteriophage:
273: sp-archaea:
274: sp-bacteria:
275: sp-fungi:
276: sp-human:
277: sp-invertebrate:
278: sp-mammal:
279: sp-mhc:
280: sp-organism:
281: sp-phage:
282: sp-plant:
283: sp-rodent:
284: sp-virus:
285: sp-vertebrate:
286: sp-unclassified:
287: sp-virus:
288: sp-bacteriophage:
289: sp-archaea:
290: sp-bacteria:
291: sp-fungi:
292: sp-human:
293: sp-invertebrate:
294: sp-mammal:
295: sp-mhc:
296: sp-organism:
297: sp-phage:
298: sp-plant:
299: sp-rodent:
300: sp-virus:
301: sp-vertebrate:
302: sp-unclassified:
303: sp-virus:
304: sp-bacteriophage:
305: sp-archaea:
306: sp-bacteria:
307: sp-fungi:
308: sp-human:
309: sp-invertebrate:
310: sp-mammal:
311: sp-mhc:
312: sp-organism:
313: sp-phage:
314: sp-plant:
315: sp-rodent:
316: sp-virus:
317: sp-vertebrate:
318: sp-unclassified:
319: sp-virus:
320: sp-bacteriophage:
321: sp-archaea:
322: sp-bacteria:
323: sp-fungi:
324: sp-human:
325: sp-invertebrate:
326: sp-mammal:
327: sp-mhc:
328: sp-organism:
329: sp-phage:
330: sp-plant:
331: sp-rodent:
332: sp-virus:
333: sp-vertebrate:
334: sp-unclassified:
335: sp-virus:
336: sp-bacteriophage:
337: sp-archaea:
338: sp-bacteria:
339: sp-fungi:
340: sp-human:
341: sp-invertebrate:
342: sp-mammal:
343: sp-mhc:
344: sp-organism:
345: sp-phage:
346: sp-plant:
347: sp-rodent:
348: sp-virus:
349: sp-vertebrate:
350: sp-unclassified:
351: sp-virus:
352: sp-bacteriophage:
353: sp-archaea:
354: sp-bacteria:
355: sp-fungi:
356: sp-human:
357: sp-invertebrate:
358: sp-mammal:
359: sp-mhc:
360: sp-organism:
361: sp-phage:
362: sp-plant:
363: sp-rodent:
364: sp-virus:
365: sp-vertebrate:
366: sp-unclassified:
367: sp-virus:
368: sp-bacteriophage:
369: sp-archaea:
370: sp-bacteria:
371: sp-fungi:
372: sp-human:
373: sp-invertebrate:
374: sp-mammal:
375: sp-mhc:
376: sp-organism:
377: sp-phage:
378: sp-plant:
379: sp-rodent:
380: sp-virus:
381: sp-vertebrate:
382: sp-unclassified:
383: sp-virus:
384: sp-bacteriophage:
385: sp-archaea:
386: sp-bacteria:
387: sp-fungi:
388: sp-human:
389: sp-invertebrate:
390: sp-mammal:
391: sp-mhc:
392: sp-organism:
393: sp-phage:
394: sp-plant:
395: sp-rodent:
396: sp-virus:
397: sp-vertebrate:
398: sp-unclassified:
399: sp-virus:
400: sp-bacteriophage:
401: sp-archaea:
402: sp-bacteria:
403: sp-fungi:
404: sp-human:
405: sp-invertebrate:
406: sp-mammal:
407: sp-mhc:
408: sp-organism:
409: sp-phage:
410: sp-plant:
411: sp-rodent:
412: sp-virus:
413: sp-vertebrate:
414: sp-unclassified:
415: sp-virus:
416: sp-bacteriophage:
417: sp-archaea:
418: sp-bacteria:
419: sp-fungi:
420: sp-human:
421: sp-invertebrate:
422: sp-mammal:
423: sp-mhc:
424: sp-organism:
425: sp-phage:
426: sp-plant:
427: sp-rodent:
428: sp-virus:
429: sp-vertebrate:
430: sp-unclassified:
431: sp-virus:
432: sp-bacteriophage:
433: sp-archaea:
434: sp-bacteria:
435: sp-fungi:
436: sp-human:
437: sp-invertebrate:
438: sp-mammal:
439: sp-mhc:
440: sp-organism:
441: sp-phage:
442: sp-plant:
443: sp-rodent:
444: sp-virus:
445: sp-vertebrate:
446: sp-unclassified:
447: sp-virus:
448: sp-bacteriophage:
449: sp-archaea:
450: sp-bacteria:
451: sp-fungi:
452: sp-human:
453: sp-invertebrate:
454: sp-mammal:
455: sp-mhc:
456: sp-organism:
457: sp-phage:
458: sp-plant:
459: sp-rodent:
460: sp-virus:
461: sp-vertebrate:
462: sp-unclassified:
463: sp-virus:
464: sp-bacteriophage:
465: sp-archaea:
466: sp-bacteria:
467: sp-fungi:
468: sp-human:
469: sp-invertebrate:
470: sp-mammal:
471: sp-mhc:
472: sp-organism:
473: sp-phage:
474: sp-plant:
475: sp-rodent:
476: sp-virus:
477: sp-vertebrate:
478: sp-unclassified:
479: sp-virus:
480: sp-bacteriophage:
481: sp-archaea:
482: sp-bacteria:
483: sp-fungi:
484: sp-human:
485: sp-invertebrate:
486: sp-mammal:
487: sp-mhc:
488: sp-organism:
489: sp-phage:
490: sp-plant:
491: sp-rodent:
492: sp-virus:
493: sp-vertebrate:
494: sp-unclassified:
495: sp-virus:
496: sp-bacteriophage:
497: sp-archaea:
498: sp-bacteria:
499: sp-fungi:
500: sp-human:
501: sp-invertebrate:
502: sp-mammal:
503: sp-mhc:
504: sp-organism:
505: sp-phage:
506: sp-plant:
507: sp-rodent:
508: sp-virus:
509: sp-vertebrate:
510: sp-unclassified:
511: sp-virus:
512: sp-bacteriophage:
513: sp-archaea:
514: sp-bacteria:
515: sp-fungi:
516: sp-human:
517: sp-invertebrate:
518: sp-mammal:
519: sp-mhc:
520: sp-organism:
521: sp-phage:
522: sp-plant:
523: sp-rodent:
524: sp-virus:
525: sp-vertebrate:
526: sp-unclassified:
527: sp-virus:
528: sp-bacteriophage:
529: sp-archaea:
530: sp-bacteria:
531: sp-fungi:
532: sp-human:
533: sp-invertebrate:
534: sp-mammal:
535: sp-mhc:
536: sp-organism:
537: sp-phage:
538: sp-plant:
539: sp-rodent:
540: sp-virus:
541: sp-vertebrate:
542: sp-unclassified:
543: sp-virus:
544: sp-bacteriophage:
545: sp-archaea:
546: sp-bacteria:
547: sp-fungi:
548: sp-human:
549: sp-invertebrate:
550: sp-mammal:
551: sp-mhc:
552: sp-organism:
553: sp-phage:
554: sp-plant:
555: sp-rodent:
556: sp-virus:
557: sp-vertebrate:
558: sp-unclassified:
559: sp-virus:
560: sp-bacteriophage:
561: sp-archaea:
562: sp-bacteria:
563: sp-fungi:
564: sp-human:
565: sp-invertebrate:
566: sp-mammal:
567: sp-mhc:
568: sp-organism:
569: sp-phage:
570: sp-plant:
571: sp-rodent:
572: sp-virus:
573: sp-vertebrate:
574: sp-unclassified:
575: sp-virus:
576: sp-bacteriophage:
577: sp-archaea:
578: sp-bacteria:
579: sp-fungi:
580: sp-human:
581: sp-invertebrate:
582: sp-mammal:
583: sp-mhc:
584: sp-organism:
585: sp-phage:
586: sp-plant:
587: sp-rodent:
588: sp-virus:
589: sp-vertebrate:
590: sp-unclassified:
591: sp-virus:
592: sp-bacteriophage:
593: sp-archaea:
594: sp-bacteria:
595: sp-fungi:
596: sp-human:
597: sp-invertebrate:
598: sp-mammal:
599: sp-mhc:
600: sp-organism:
601: sp-phage:
602: sp-plant:
603: sp-rodent:
604: sp-virus:
605: sp-vertebrate:
606: sp-unclassified:
607: sp-virus:
608: sp-bacteriophage:
609: sp-archaea:
610: sp-bacteria:
611: sp-fungi:
612: sp-human:
613: sp-invertebrate:
614: sp-mammal:
615: sp-mhc:
616: sp-organism:
617: sp-phage:
618: sp-plant:
619: sp-rodent:
620: sp-virus:
621: sp-vertebrate:
622: sp-unclassified:
623: sp-virus:
624: sp-bacteriophage:
625: sp-archaea:
626: sp-bacteria:
627: sp-fungi:
628: sp-human:
629: sp-invertebrate:
630: sp-mammal:
631: sp-mhc:
632: sp-organism:
633: sp-phage:
634: sp-plant:
635: sp-rodent:
636: sp-virus:
637: sp-vertebrate:
638: sp-unclassified:
639: sp-virus:
640: sp-bacteriophage:
641: sp-archaea:
642: sp-bacteria:
643: sp-fungi:
644: sp-human:
645: sp-invertebrate:
646: sp-mammal:
647: sp-mhc:
648: sp-organism:
649: sp-phage:
650: sp-plant:
651: sp-rodent:
652: sp-virus:
653: sp-vertebrate:
654: sp-unclassified:
655: sp-virus:
656: sp-bacteriophage:
657: sp-archaea:
658: sp-bacteria:
659: sp-fungi:
660: sp-human:
661: sp-invertebrate:
662: sp-mammal:
663: sp-mhc:
664: sp-organism:
665: sp-phage:
666: sp-plant:
667: sp-rodent:
668: sp-virus:
669: sp-vertebrate:
670: sp-unclassified:
671: sp-virus:
672: sp-bacteriophage:
673: sp-archaea:
674: sp-bacteria:
675: sp-fungi:
676: sp-human:
677: sp-invertebrate:
678: sp-mammal:
679: sp-mhc:
680: sp-organism:
681: sp-phage:
682: sp-plant:
683: sp-rodent:
684: sp-virus:
685: sp-vertebrate:
686: sp-unclassified:
687: sp-virus:
688: sp-bacteriophage:
689: sp-archaea:
690: sp-bacteria:
691: sp-fungi:
692: sp-human:
693: sp-invertebrate:
694: sp-mammal:
695: sp-mhc:
696: sp-organism:
697: sp-phage:
698: sp-plant:
699: sp-rodent:
700: sp-virus:
701: sp-vertebrate:
702: sp-unclassified:
703: sp-virus:
704: sp-bacteriophage:
705: sp-archaea:
706: sp-bacteria:
707: sp-fungi:
708: sp-human:
709: sp-invertebrate:
710: sp-mammal:
711: sp-mhc:
712: sp-organism:
713: sp-phage:
714: sp-plant:
715: sp-rodent:
716: sp-virus:
717: sp-vertebrate:
718: sp-unclassified:
719: sp-virus:
720: sp-bacteriophage:
721: sp-archaea:
722: sp-bacteria:
723: sp-fungi:
724: sp-human:
725: sp-invertebrate:
726: sp-mammal:
727: sp-mhc:
728: sp-organism:
729: sp-phage:
730: sp-plant:
731: sp-rodent:
732: sp-virus:
733: sp-vertebrate:
734: sp-unclassified:
735: sp-virus:
736: sp-bacteriophage:
737: sp-archaea:
738: sp-bacteria:
739: sp-fungi:
740: sp-human:
741: sp-invertebrate:
742: sp-mammal:
743: sp-mhc:
744: sp-organism:
745: sp-phage:
746: sp-plant:
747: sp-rodent:
748: sp-virus:
749: sp-vertebrate:
750: sp-unclassified:
751: sp-virus:
752: sp-bacteriophage:
753: sp-archaea:
754: sp-bacteria:
755: sp-fungi:
756: sp-human:
757: sp-invertebrate:
758: sp-mammal:
759: sp-mhc:
760: sp-organism:
761: sp-phage:
762: sp-plant:
763: sp-rodent:
764: sp-virus:
765: sp-vertebrate:
766: sp-unclassified:
767: sp-virus:
768: sp-bacteriophage:
769: sp-archaea:
770: sp-bacteria:
771: sp-fungi:
772: sp-human:
773: sp-invertebrate:
774: sp-mammal:
775: sp-mhc:
776: sp-organism:
777: sp-phage:
778: sp-plant:
779: sp-rodent:
780: sp-virus:
781: sp-vertebrate:
782: sp-unclassified:
783: sp-virus:
784: sp-bacteriophage:
785: sp-archaea:
786: sp-bacteria:
787: sp-fungi:
788: sp-human:
789: sp-invertebrate:
790: sp-mammal:
791: sp-mhc:
792: sp-organism:
793: sp-phage:
794: sp-plant:
795: sp-rodent:
796: sp-virus:
797: sp-vertebrate:
798: sp-unclassified:
799: sp-virus:
800: sp-bacteriophage:
801: sp-archaea:
802: sp-bacteria:
803: sp-fungi:
804: sp-human:
805: sp-invertebrate:
806: sp-mammal:
807: sp-mhc:
808: sp-organism:
809: sp-phage:
810: sp-plant:
811: sp-rodent:
812: sp-virus:
813: sp-vertebrate:
814: sp-unclassified:
815: sp-virus:
816: sp-bacteriophage:
817: sp-archaea:
818: sp-bacteria:
819: sp-fungi:
820: sp-human:
821: sp-invertebrate:
822: sp-mammal:
823: sp-mhc:
824: sp-organism:
825: sp-phage:
826: sp-plant:
827: sp-rodent:
828: sp-virus:
829: sp-vertebrate:
830: sp-unclassified:
831: sp-virus:
832: sp-bacteriophage:
833: sp-archaea:
834: sp-bacteria:
835: sp-fungi:
836: sp-human:
837: sp-invertebrate:
838: sp-mammal:
839: sp-mhc:
840: sp-organism:
841: sp-phage:
842: sp-plant:
843: sp-rodent:
844: sp-virus:
845: sp-vertebrate:
846: sp-unclassified:
847: sp-virus:
848: sp-bacteriophage:
849: sp-archaea:
850: sp-bacteria:
851: sp-fungi:
852: sp-human:
853: sp-invertebrate:
854: sp-mammal:
855: sp-mhc:
856: sp-organism:
857: sp-phage:
858: sp-plant:
859: sp-rodent:
860: sp-virus:
861: sp-vertebrate:
862: sp-unclassified:
863: sp-virus:
864: sp-bacteriophage:
865: sp-archaea:
866: sp-bacteria:
867: sp-fungi:
868: sp-human:
869: sp-invertebrate:
870: sp-mammal:
871: sp-mhc:
872: sp-organism:
873: sp-phage:
874: sp-plant:
875: sp-rodent:
876: sp-virus:
877: sp-vertebrate:
878: sp-unclassified:
879: sp-virus:
880: sp-bacteriophage:
881: sp-archaea:
882: sp-bacteria:
883: sp-fungi:
884: sp-human:
885: sp-invertebrate:
886: sp-mammal:
887: sp-mhc:
888: sp-organism:
889: sp-phage:
890: sp-plant:
891: sp-rodent:
892: sp-virus:
893: sp-vertebrate:
894: sp-unclassified:
895: sp-virus:
896: sp-bacteriophage:
897: sp-archaea:
898: sp-bacteria:
899: sp-fungi:
900: sp-human:
901: sp-invertebrate:
902: sp-mammal:
903: sp-mhc:
904: sp-organism:
905: sp-phage:
906: sp-plant:
907: sp-rodent:
908: sp-virus:
909: sp-vertebrate:
910: sp-unclassified:
911: sp-virus:
912: sp-bacteriophage:
913: sp-archaea:
914: sp-bacteria:
915: sp-fungi:
916: sp-human:
917: sp-invertebrate:
918: sp-mammal:
919: sp-mhc:
920: sp-organism:
921: sp-phage:
922: sp-plant:
923: sp-rodent:
924: sp-virus:
925: sp-vertebrate:
926: sp-unclassified:
927: sp-virus:
928: sp-bacteriophage:
929: sp-archaea:
930: sp-bacteria:
931: sp-fungi:
932: sp-human:
933: sp-invertebrate:
934: sp-mammal:
935: sp-mhc:
936: sp-organism:
937: sp-phage:
938: sp-plant:
939: sp-rodent:
940: sp-virus:
941: sp-vertebrate:
942: sp-unclassified:
943: sp-virus:
944: sp-bacteriophage:
945: sp-archaea:
946: sp-bacteria:
947: sp-fungi:
948: sp-human:
949: sp-invertebrate:
950: sp-mammal:
951: sp-mhc:
952: sp-organism:
953: sp-phage:
954: sp-plant:
955: sp-rodent:
956: sp-virus:
957: sp-vertebrate:
958: sp-unclassified:
959: sp-virus:
960: sp-bacteriophage:
961: sp-archaea:
962: sp-bacteria:
963: sp-fungi:
964: sp-human:
965: sp-invertebrate:
966: sp-mammal:
967: sp-mhc:
968: sp-organism:
969: sp-phage:
970: sp-plant:
971: sp-rodent:
972: sp-virus:
973: sp-vertebrate:
974: sp-unclassified:
975: sp-virus:
976: sp-bacteriophage:
977: sp-archaea:
978: sp-bacteria:
979: sp-fungi:
980: sp-human:
981: sp-invertebrate:
982: sp-mammal:
983: sp-mhc:
984: sp-organism:
985: sp-phage:
986: sp-plant:
987: sp-rodent:
988: sp-virus:
989: sp-vertebrate:
990: sp-unclassified:
991: sp-virus:
992: sp-bacteriophage:
993: sp-archaea:
994: sp-bacteria:
995: sp-fungi:
996: sp-human:
997: sp-invertebrate:
998: sp-mammal:
999: sp-mhc:
1000: sp-organism:
1001: sp-phage:
1002: sp-plant:
1003: sp-rodent:
1004: sp-virus:
1005: sp-vertebrate:
1006: sp-unclassified:
1007: sp-virus:
1008: sp-bacteriophage:
1009: sp-archaea:
1010: sp-bacteria:
1011: sp-fungi:
1012: sp-human:
1013: sp-invertebrate:
1014: sp-mammal:
1015: sp-mhc:
1016: sp-organism:
1017: sp-phage:
1018: sp-plant:
1019: sp-rodent:
1020: sp-virus:
1021: sp-vertebrate:
1022: sp-unclassified:
1023: sp-virus:
1024: sp-bacteriophage:
1025: sp-archaea:
1026: sp-bacteria:
1027: sp-fungi:
1028: sp-human:
1029: sp-invertebrate:
1030: sp-mammal:
1031: sp-mhc:
1032: sp-organism:
1033: sp-phage:
1034: sp-plant:
1035: sp-rodent:
1036: sp-virus:
1037: sp-vertebrate:
1038: sp-unclassified:
1039: sp-virus:
1040: sp-bacteriophage:
1041: sp-archaea:
1042: sp-bacteria:
1043: sp-fungi:
1044: sp-human:
1045: sp-invertebrate:
1046: sp-mammal:
1047: sp-mhc:
1048: sp-organism:
1049: sp-phage:
1050: sp-plant:
1051: sp-rodent:
1052: sp-virus:
1053: sp-vertebrate:
1054: sp-unclassified:
1055: sp-virus:
1056: sp-bacteriophage:
1057: sp-archaea:
1058: sp-bacteria:
1059: sp-fungi:
1060: sp-human:
1061: sp-invertebrate:
1062: sp-mammal:
1063: sp-mhc:
1064: sp-organism:
1065: sp-phage:
1066: sp-plant:
1067: sp-rodent:
1068: sp-virus:
1069: sp-vertebrate:
1070: sp-unclassified:
1071: sp-virus:
1072: sp-bacteriophage:
1073: sp-archaea:
1074: sp-bacteria:
1075: sp-fungi:
1076: sp-human:
1077: sp-invertebrate:
1078: sp-mammal:
1079: sp-mhc:
1080: sp-organism:
1081: sp-phage:
1082: sp-plant:
1083: sp-rodent:
1084: sp-virus:
1085: sp-vertebrate:
1086: sp-unclassified:
1087: sp-virus:
1088: sp-bacteriophage:
1089: sp-archaea:
1090: sp-bacteria:
1091: sp-fungi:
1092: sp-human:
1093: sp-invertebrate:
1094: sp-mammal:
1095: sp-mhc:
1096: sp-organism:
1097: sp-phage:
1098: sp-plant:
1099: sp-rodent:
1100: sp-virus:
1101: sp-vertebrate:
1102: sp-unclassified:
1103: sp

01 MAR 2002 (Rel. 41, last annotation update)
 DE Probable UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14) (UDP-
 GlcNAc 2-epimerase).
 GN EPSC OR RSP1017 OR R502355.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 O1 Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OR Ralstonia.
 OX NCB1 TaxID: 405;
 EN NCB1 TaxID: 405;
 RP STRAIN: SM11300;
 RC STRAIN: SM11300;
 RA Sankarabhat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Astar M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Chuisne N., Claudel-Renard C., Cunne S., Demange N.,
 RA Gaspin C., Lavié M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siquier P., Thebaud P., Whalen P., Wincker P., Levy M.,
 RA Weissbach J., Houcher C.A.
 RA Genome sequence of the plant pathogen Ralstonia solanacearum.
 RL Submitted (08/2001) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN SYNTHESIS OF N-
 ACETYLIDEXYGLUCOSE, A COMPONENT OF EXOPOLYSACCHARIDE EPS I
 WHICH FUNCTIONS AS A VIRULENCE FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine -> UDP-N-acetyl-D-
 mannosamine.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE EPS I SYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE UDP-N-ACETYLGLUCOSAMINE 2-EPIMERASE
 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/geneuse/>
 or send an email to license@sib-sib.ch).
 DR EMBL: AL446042; CA018168.1; ALT_INIT.
 DR Lipopolysaccharide biosynthesis; Isomerase; Plasmid;
 KW Complete proteome.
 SQ SEQUENCE 475 AA: 40959 MW: 43884 35602H2AR0 CRC64;

Query Match 3.5%, Score 7; DB 1; Length 375;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 59 TLATTLA 65
 I I I I I I
 I 100 TLATTLA 106

RESULT 15
 EPSC_RALS)
 ID EPSC_RALS) STANIAR: PRI: 375 AA.
 AC P52641;
 DT 01-OCT 1996 (Rel. 41, Created)
 DT 01 MAR 2002 (Rel. 41, Last sequence update)
 DT 01 MAR 2002 (Rel. 41, Last annotation update)
 DE Probable UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14) (UDP-
 GlcNAc 2-epimerase).
 GN EPSC.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OR Ralstonia.
 OX NCB1 TaxID: 405;
 EN NCB1 TaxID: 405;
 RP STRAIN: SM11300;
 RC STRAIN: SM11300;
 RA Sankarabhat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Astar M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Chuisne N., Claudel-Renard C., Cunne S., Demange N.,
 RA Gaspin C., Lavié M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siquier P., Thebaud P., Whalen P., Wincker P., Levy M.,
 RA Weissbach J., Houcher C.A.
 RA Genome sequence of the plant pathogen Ralstonia solanacearum.
 RL Submitted (08/2001) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN SYNTHESIS OF N-
 ACETYLIDEXYGLUCOSE, A COMPONENT OF EXOPOLYSACCHARIDE EPS I
 WHICH FUNCTIONS AS A VIRULENCE FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine -> UDP-N-acetyl-D-
 mannosamine.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE EPS I SYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE UDP-N-ACETYLGLUCOSAMINE 2-EPIMERASE
 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/geneuse/>
 or send an email to license@sib-sib.ch).
 DR EMBL: AL446042; CA018168.1; ALT_INIT.
 DR Lipopolysaccharide biosynthesis; Isomerase; Plasmid;
 KW Complete proteome.
 SQ SEQUENCE 475 AA: 40959 MW: 43884 35602H2AR0 CRC64;

RP solanacearum and its transcriptional regulation at a single
 RT promoter.";
 RL Mol. Microbiol. 16:977-989(1995).
 CC -1- FUNCTION: MAY BE INVOLVED IN SYNTHESIS OF N-
 ACETYLIDEXYGLUCOSE, A COMPONENT OF EXOPOLYSACCHARIDE EPS I
 WHICH FUNCTIONS AS A VIRULENCE FACTOR.
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine -> UDP-N-acetyl-D-
 mannosamine.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE EPS I SYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE UDP-N-ACETYLGLUCOSAMINE 2-EPIMERASE
 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/geneuse/>
 or send an email to license@sib-sib.ch).
 DR EMBL: U17898; AAA91626.1; ALT_INIT.
 DR HSSP: P27828; IP6D.
 DR InterPro: IPR003331; Epimerase_2.
 DR Pfam: PF02350; Epimerase_2_1.
 DR Lipopolysaccharide biosynthesis; Isomerase.
 KW SEQUENCE 375 AA: 40944 MW: B07B54DA;90F789 CRC64;

Query Match 3.5%, Score 7; DB 1; Length 375;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 59 TLATTLA 65
 I I I I I I
 I 100 TLATTLA 106

Search completed: July 2, 2002, 16:19:06
 Job time: 778 sec

*Molecular characterization of the eps gene cluster of Pseudomonas

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AF001094; AAB9163.1;
F1: AF0172;
KW: Hypothetical protein; transmembrane; Complete proteome.
F1: TRANSMEM 13 6 POTENTIAL.
F1: TRANSMEM 50 62 POTENTIAL.
F1: TRANSMEM 75 94 POTENTIAL.
F1: TRANSMEM 109 141 POTENTIAL.
F1: TRANSMEM 152 174 POTENTIAL.
F1: TRANSMEM 194 216 POTENTIAL.
SQ: SEQUENCE 217 AA; 24847 MW; 96148F0C267C06 CRC64;

Query Match 3.5% Score 7; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAA 10
DB 5 LLSALAA 64

RESULT 11
FPG_PASMO
ID FPG_PASMO STANDARD: PRT: 213 AA.
AC 084716;
DI 15 DEC-1998 (Rel. 47, Last sequence update)
DI 16 DEC-2001 (Rel. 40, Last annotation update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (inosine phosphorylase)
DE (PNP).
GN DEPO OR TP0744.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID:160;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN: NICHOLS;
PX MEDLINE-98432770; PubMed-9665876;
KA Flaser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
KA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
KA Soderstrom E., Hardham T.M., McLeod M.P., Salzberg S., Peterson T.,
KA Khalak H., Richardson D., Howell I.K., Chidambaram M., Hirschback T.,
KA McDonald L., Artach P., Bowman G., Cohen M.D., Fujii C., Garland S.,
KA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
KA Venter J.C.;
PI Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
RE Science 281:375-382(1998)
CC FUNCTION: CLEAVAGE OF GUANOSINE OR INOSINE TO RESPECTIVE BASES AND
CC SUGAR 1 PHOSPHATE MOieties (BY SIMILARITY)
CC -1- CATALYTIC ACTIVITY: Purine nucleoside + phosphate -> purine +
CC alpha-D-ribose 1-phosphate.
CC -1- SUBUNIT: HOMHEXAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PNP/ODP FAMILY 1 OF PHOSPHORYLASES
CC THIS SWISS PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: AF001245; AA065702.1;
DB HSP: TP0744; IAB9.
DB TIGR: TP0744;

DR InterPro: IPR000845; PNP_ODP.
DR Pfam: PF01048; PNP_ODP_1; 1.
DR PROSITE: PS01232; PNP_ODP_1; 1.
KW Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 233 AA; 25318 MW; DIA94487H744A23 CRC64;

Query Match 3.5% Score 7; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 APRVLLP 71
DB 13 APRVLLP 19

RESULT 11
FPG_PASMO
ID FPG_PASMO STANDARD: PRT: 270 AA.
AC P57910;
DI 16 OCT-2001 (Rel. 40, Created)
DI 16 OCT-2001 (Rel. 40, Last sequence update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.2) (Fapy DNA
DE glycosylase).
GN MTHM OR FPG OR PM1145.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID-747;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-PM70;
PX MEDLINE-21145866; PubMed-11248100;
KA May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genome sequence of Pasteurella multocida Em70.1";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001)
CC -1- FUNCTION: THIS ENZYME MAY PLAY A SIGNIFICANT ROLE IN PROCESSES
CC LEADING TO RECOVERY FROM MUTAGENESIS AND/OR CELL DEATH BY
CC ALKYLATING AGENTS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N7
CC methylguanine residues, releasing 2,6-diamino 4-hydroxy-5-(N
CC methyl)formamidopyrimidine.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FPG FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AE006155; AAK03229.1;
DR InterPro: IPR000191; Fapy_DNA_glyco.
DR InterPro: IPR000214; Fapy_DNA_glyco_2n.
DR Pfam: PF01149; Fapy_DNA_glyco_1.
DR ProDom: PD003680; Fapy_DNA_glyco_1.
DR PROSITE: PS01242; FPG; 1.
KW DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
KW Complete proteome.
FI ZN-FING 244 267 POTENTIAL.
SQ SEQUENCE 270 AA; 30612 MW; B44C3B0F0D968734 CRC64;

Query Match 3.5% Score 7; DB 1; Length 270;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ISPVLEG 28
DB 14 ISPVLEG 20

Gregor J., Davis N.W., Kirkpatrick H.A., Gordon M.A., Poso R.J.,
 Mau B., Shao Y.;
 RA "The complete genome sequence of *Escherichia coli* K-12.";
 SC Science 277:1453-1474 (1997).
 [4]
 RP SEQUENCE FROM N.A.
 EX MEDLINE:947251 (6); PubMed:9497940;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Ikeda T.
 PA Isono K., Kasai H., Kimura S., Kikakawa M., Kikakawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakano S., Nakamura Y., Nishimoto H., Nishio Y., Oshima H.,
 RA Saito N., Sempel G., Seki Y., Sivassundaram S., Tajima H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;
 RA "A 460 kb DNA sequence of the *Escherichia coli* K-12 genome
 PT corresponding to the 40.1-50.0 min region on the linkage map.";
 RA DNA Res. 3:479-492 (1996).
 [5]
 RP SEQUENCE FROM N.A.
 EX STRAIN:0157:H7 / EDL 943 / A/C07 7009427;
 RA MEDLINE:21074945; PubMed:11296561;
 PA Berna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Groback E.J., Davis N.W., Lim A., Bimalanta E.F., Rocamousta K.,
 RA Apodaca J., Anantharaman F.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RA "Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.";
 RA Nature 409:529-533 (2001).
 [6]
 RP SEQUENCE FROM N.A.
 EX STRAIN:0157:H7 / PMID 05099952;
 RA MEDLINE:21156241; PubMed:11258796;
 PA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 PT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RA DNA Res. 8:11-22 (2001).
 [7]
 RP FUNCTION: RCSP IS A MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM.
 CC RCSP/RCSC, WHICH REGULATES THE EXPRESSION OF GENES INVOLVED IN
 CC COLANIC ACID CAPSULE SYNTHESIS, RCSP ACTS AS THE EFFECTOR.
 CC ADDITIONAL CONTROL IS PROVIDED BY THE DEPENDENCE ON THE ALTERNATE
 CC SIGMA FACTOR, RPN, FOR THE SYNTHESIS OF RCSP, RCSC AND RCSP FORM
 CC A COMPLEX TO PROMOTE TRANSCRIPTION OF THE GENES FOR CAPSULE
 CC SYNTHESIS.
 [8]
 RP FUNCTION: ALSO STIMULATES FCSP EXPRESSION. THIS SUGGESTS THAT RCSP
 CC IS A REGULATOR IMPLICATED IN THE REGULATION OF MORE THAN ONE
 CC CELLULAR FUNCTION.
 CC SIMILARITY: BELONGS TO THE TOX-2/TOXA FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 [9]
 RP SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
 [10]
 RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annoucy/>
 CC or send an email to license@sib-sib.ch).
 [11]
 RA EMBL: M28242; AAA24504.1;
 RA EMBL: L11272; AAA24506.1;
 RA EMBL: AF000410; AAC75277.1;
 RA EMBL: D90850; BAA16000.1;
 RA EMBL: D90851; BAA16008.1;
 RA EMBL: AF005453; AAG57452.1;
 RA EMBL: AF002560; BAB46529.1;
 RA PIR: J00068; RVE0CB
 RA EMBL: EG10821; RCSP
 RA EMBL: U0000792; HTH_LoxR
 RA InterPro: I0001789; Response_req.

PFam: PF00196; GcRE: 1;
 PFam: PF00072; Response_req: 1;
 DR PRINTS: PR00038; HTH_LoxR: 1;
 DR SMART: SM00421; HTH_LoxR: 1;
 DR SMART: SM00448; REC: 1;
 DR PROSITE: PS00622; HTH_LoxR_FAMILY: 1;
 DR PROSITE: PS0110; RESPONSE_REGULATORY: 1;
 KW Sensory transduction; Phosphorylation; Transcription regulation;
 FW DNA-binding; Activator; Complete proteome.
 FT DOMAIN 1 124 RESPONSE REGULATORY
 FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY);
 FT DNA_BIND 168 187 H-T-H MOTIF (BY SIMILARITY);
 SQ SEQUENCE 216 AA; 74670 MW; A78D030344F96B9 CR64;

Query Match 3.5%; Score 7; DB 1; Length 216;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ALAALQK 13
 Db 119 ALAALQK 125

RESUL 7
 ID RCSP_SALTI STANDARD; PRT: 216 AA.
 AC Q56127;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Capsular synthesis regulator component B.

GN RCSP OR STY2495.

OS Salmonella typhi.

OC Bacteria; Proteobacteria, gamma subdivision, Enterobacteriaceae;

OC Salmonella.

OX NCBI_TaxID:601;

RP SEQUENCE FROM N.A.

RC STRAIN-TY2;

EX MEDLINE 96198173; PubMed-8426298;

RA Virlogeux F., Waxin H., Ecobichon C., Lee J.-O., Popoff M.Y.;

PT "Characterization of the *tcsA* and *rcsp* genes from *Salmonella typhi*;

PT *rcsp* through *tvaA* is involved in regulation of Vi antigen

PT synthesis.";

RL J. Pathol. 178:1691-1696 (1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-GIFU 10007;

RA Hashimoto Y.;

PL Submitted (5/9-1997) to the EMBL/GenBank/DBJ databases.

PN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-CT18;

EX MEDLINE-21534947; PubMed-11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.K., Parkard D., Wain J.,

PA Churcher C., Morgan F.L., Bentley S.D., Holden M.T.G., Seakula M.,

Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

PA Feltwell T., Hamlin N., Hance A., Hien T.T., Holroyd S., Jancels K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,

RA Quail M., Rutherford K., Simmonds M., Skellern J., Stevens K.,

RA Whitehead S., Barrell B.G.;

PT "Complete genome sequence of a multiple drug resistant *Salmonella*

PT enterica serovar Typhi CT18.";

RL Nature 413:848-852 (2001).

CC FUNCTION: RCSP IS A MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM,

CC RCSP/RCSC, WHICH REGULATES THE EXPRESSION OF GENES INVOLVED IN

CC COLANIC ACID CAPSULE SYNTHESIS, RCSP ACTS AS THE EFFECTOR.

CC ADDITIONAL CONTROL IS PROVIDED BY THE DEPENDENCE ON THE ALTERNATE

CC SIGMA FACTOR, RPN, FOR THE SYNTHESIS OF RCSP, RCSC AND RCSP FORM

CC A COMPLEX TO PROMOTE TRANSCRIPTION OF THE GENES FOR CAPSULE

CC SYNTHESIS.

Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: MAY PLAY A ROLE IN L-LACTATE TRANSPORT.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE LIDP FAMILY OF TRANSPORTERS
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z94043; CA080902.1; -;
 DR EMBL: Z71928; CAA96486.1; -;
 DR EMBL: Z99121; CAB15424.1; -;
 DR Subtilist; BG11875; yviih.
 DR InterPro: IPR003804; lactate_perm.
 DR Pfam: PF02652; lactate_perm; 1.
 KW Hypothetical protein, Transpositt, Transmembrane, Complete proteome.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 37 57 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 220 240 POTENTIAL.
 FT TRANSMEM 249 269 POTENTIAL.
 FT TRANSMEM 304 324 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 FT TRANSMEM 448 468 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 542 562 POTENTIAL.
 SQ SEQUENCE 563 AA; 59761 MW; E380983059B19D08 CRC64;

Query Match 4.0%; Score 8; DB 1; Length 563;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LISAIAAL 11
 DB 15 LISAIAAL 22

RESULT 5
 YRAW_ECOLI STANDARD; DBT; 132 AA
 AC Y77712;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yraw.
 GN YRAW OR H0443
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Peto J.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K 12";
 RT Science 277:1453-1474(1997).
 RN 12
 RP SEQUENCE FROM N.A.
 RC ROBERTS D., Allon R., Araujo R., Aparicio A., Chung H., Davis K.,
 RA Duncan M., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,

RA Icw H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
 RI Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
 RN 13
 RP SEQUENCE FROM N.A
 RC STRAIN-K12 / W3110;
 RA Hatada F., Ohmori H., Qiao Y., Tsuji M., Fukuda R.;
 RI Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000150; AAC73546.1; -;
 DR EMBL: DR2664; AAB40199.1; -;
 DR EMBL: DR2943; AAB11647.1; -;
 DR ECGGene; E31325.1; ybaw.
 DR InterPro: IPR000365; 4HbcoA_thioester.
 DR Pfam: PF03061; 4HBT; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 132 AA; 15088 MW; FBH3539149A37383 CRC64;

Query Match 3.5%; Score 7; DB 1; Length 132;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 GILSQVI 34
 DB 84 GILSQVI 90

RESULT 6
 RCSB_ECOLI STANDARD; DBT; 216 AA.
 AC P14374;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Capsular synthesis regulator component H.
 GN RCPH OR H2217 OR P3476 OR PCS3106
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-90130299; PubMed-2404948;
 RA Stout V., Gottesman S.;
 RA "RcsB and RcsC: a two-component regulator of capsule synthesis in
 RA Escherichia coli";
 RT J. Bacteriol. 172:659-669(1990).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=09:K30:H12;
 RX MEDLINE-93374832; PubMed-8366025;
 RA Jayaratne P., Keenleyside W.J., MacLachlan P.R., Dodgson C.,
 RA Whitfield C.;
 RA "Characterization of rcsB and rcsC from Escherichia coli O9:K30:H12
 RA and examination of the role of the rcs regulatory system in
 RA expression of group 1 capsular polysaccharides";
 RL J. Bacteriol. 175:5384-5394(1993).
 RN 13
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Peto J.K., Mayhew G.F.,

```

F1 VARIANI 2017 2013 /ET14-VAR_010941.
F1 E - G.
F1 /ET14-VAR_010942.
S1 SEQUENCE 2144 AA: 42455 MW: D66816EE780BC9B7 CRC64:

Query Match
Best Local Similarity 100.0%; Score 100; DR 1; Length 2144;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

97 1 EVYLLSALAAQGVVFIPHEISYVIFHLSAVHLEFVTFSEMSASQANTPIFSGVFTL 60
DB 1754 EVYLLSALAAQGVVFIPHEISYVIFHLSAVHLEFVTFSEMSASQANTPIFSGVFTL 1813
QY 61 ATTLPAPVITPAIKTKYQIFKKNKNNM:PFMSITQEHIG 100
DB 1814 ATTLPAPVITPAIKTKYQIFKKNKNNM:PFMSITQEHIG 1853

RESULT 4
RP28_MATEA STANDARD: PRT: 958 AA.
AC Q9GMA4:
DI 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein BAP28 (Fragment).
GN BAP28.
OS Macaca fascicularis (Grab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
KW TISSUE=Brain;
RA Osada N., Hida M., Kusuda J., Ianuma K., Iseki K., Hirai M., Ierao K.,
RA Suzuki Y., Sudano S., Hashimoto K.
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT Libraries."
RL Submitted (OCT 2000) to the EMBL/GenBank/TrEMBL databases.
CC -!- SIMILARITY: BELONGS TO THE BAP28 FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HEAT REPEAT.
CC This SWISS-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
DB EMBL: AB049842; BAB16728.1; ALT_INIT.
DB TrEMBL: IP4000457; HEAT_REPEAT.
DB PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.
F1 NON-TER 1
F1 REPEAT 920 956 HEAT.
SQ SEQUENCE 958 AA: 108644 MW: 438095536230FB31 CRC64:

Query Match
Best Local Similarity 100.0%; Score 63; DR 1; Length 958;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

97 149 TENITDCLVAMVKLSVTFPPFPEKLEFWAKTEADPKDPLIFNYLADYIAEKIKGLE 198
DB 706 TENCITDCLVAMVKLSVTFPPFPEKLEFWAKTEADPKDPLIFNYLADYIAEKIKGLE 765
QY 199 TLF 201
DB 766 TLF 768

```

```

RESULT 3
YA97_MYCPN STANDARD: PRT: 541 AA.
AC P75595:
DI 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein MPN097 precursor (M02_ort541).
GN MPN097 OR MP057.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
EX MEDLINE=97105885; PubMed=8948683;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel K., Li H.-C.,
RA Herrmann K.
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE MCL15 / MCL260 FAMILY.
CC This SWISS-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
DB EMBL: AE000007; AAB95705.1;
KW Hypothetical protein, lipoprotein; Membrane; Signal;
KW Complete proteome.
F1 SIGNAL 1 22 POTENTIAL.
F1 CHAIN 23 541
F1 LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 541 AA: 59153 MW: A037CB2B826B1E9C CRC64:

Query Match
Best Local Similarity 100.0%; Score 8; DR 1; Length 541;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AALQKVE 16
DB 50 AALQKVE 57

RESULT 4
YVPH_HACSU STANDARD: PRT: 563 AA.
AC P71067:
DI 01-NOV-1997 (Rel. 35, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative L lactate permease YVPH.
GN YVPH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Denizot F.C.;
RL Submitted (APR 1997) to the EMBL/GenBank/TrEMBL databases.
RN [2]
RP SEQUENCE OF 51-563 FROM N.A.
RC STRAIN=168;
RA Fabret C., Quentin Y., Chapal N., Guisepi A., Haiech J., Denizot F.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:19:06, Search time: 40.19 seconds
(without alignments)
257.788 Million cell updates/sec

Title: US-09-603-665-5_COPY_1754_1954

Perfect score: 201

Sequence: 1 EVYLLSALAALQKWTLPH TFYNLADCTAEKLGKLTFLF 201

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size: 6

Total number of hits satisfying chosen parameters: 463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	49.8	2144	1	BP28_HUMAN
2	63	31.3	958	1	BP28_MACFA
3	8	4.0	541	1	YA97_MYCPN
4	8	4.0	563	1	YVFL_BACSU
5	7	3.5	132	1	YHAM_ECOLI
6	7	3.5	216	1	RCSB_ECOLI
7	7	3.5	216	1	RCSB_SALTI
8	7	3.5	217	1	Y172_ARCFU
9	7	3.5	233	1	DEOD_TPEPA
10	7	3.5	270	1	FINC_PASMU
11	7	3.5	311	1	OB02_HUMAN
12	7	3.5	358	1	PIAP_PIG
13	7	3.5	375	1	EPCL_RALSO
14	7	3.5	375	1	EPCL_RALSO
15	7	3.5	385	1	YB01_MYCTU
16	7	3.5	414	1	YQJH_RACSU
17	7	3.5	440	1	Y788_BOHBU
18	7	3.5	476	1	MM10_RAT
19	7	3.5	478	1	NRIA_ECOLI
20	7	3.5	484	1	AMYA_ASPNG
21	7	3.5	537	1	CYEA_AZSVI
22	7	3.5	550	1	MP10_XENIA
23	7	3.5	550	1	MP13_XENIA
24	7	3.5	572	1	MP13_XENIA
25	7	3.5	579	1	IMDH_MAIZE
26	7	3.5	599	1	MP12_XENIA
27	7	3.5	709	1	GATB_HUMAN
28	7	3.5	794	1	YH52_MYCPN
29	7	3.5	883	1	CAPP_ECOLI
30	7	3.5	960	1	L136_CAFEL
31	7	3.5	960	1	L136_CAFEL
32	7	3.5	1325	1	YAE6_SCHYO
33	7	3.5	1434	1	PTC1_MOUSE

RESULT 1

ID BP28_HUMAN STANDARD: PRT: 2144 AA.
AC Q9H583; Q9NW23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein BAP28.
GN BAP28.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND VARIANTS S-1694; A-1854; D-1967 AND G-2017.
RA Bougueleret L., Ghumkov I., Barry C., Cohen Akenine A.
RT "A novel BAP28 gene and protein."
RL Patent number WO0100669, 04-JAN-2001.
RN [2]
RP SEQUENCE OF 1534-2144 FROM N.A.
RA Cobley V.;
RC Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1777-2144 FROM N.A.
RA Isoqai T., Ota I., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Suqano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Nakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "New human cDNA sequencing project";
FL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC 1 SIMILARITY: RELIGIONS TO THE BAP28 FAMILY.
CC 1 SIMILARITY: CONTAINS 1 HEAT REPEAT.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (see <http://www.isb.sib.ch/announce/> or send an email to license@sib.sib.ch)

EMBL: AX067150; CAC26776.1;
EMBL: A1136105; CAC15948.1;
EMBL: AK01221; BAA91564.1; ALT_INIT.
InterPro: IPR000357, HEAT_REPEAT.
PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.
Polymorphism: 2143
PEPVAR 2143
VARIANT 1694
VARIANT 1854
VARIANT 1967
VARIANT 1967

34 7 3.5 1442 1 PTC1_CHICK
35 7 3.5 1447 1 PTC1_HUMAN
36 7 3.5 1453 1 Y373_BOVIN
37 7 3.5 1539 1 Y373_HUMAN
38 7 3.5 1612 1 ATC4_YEAST
39 6 3.0 35 1 RL7_BUCAP
40 6 3.0 40 1 PSAL_CYAPA
41 6 3.0 62 1 YM45_CAEEL
42 6 3.0 66 1 XEN2_XENIA
43 6 3.0 66 1 XEN3_XENIA
44 6 3.0 68 1 YF02_HAELIN
45 6 3.0 69 1 YAT1_RHORD

ALIGNMENTS

Q90693 gallus gall
Q13635 homo sapien
Q9U223 bos taurus
O15078 homo sapien
Q12675 saccharomyc
P41188 buchura ap
P48117 cyanophora
P34521 caenorhabdi
P38951 xenopus lae
P38952 xenopus lae
P44177 haemophilus
P05443 rhodospseudo

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

A:Accession: A44909
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 4-9 <GR>
A:Cross-references: GR.S37760, NID.q250030, PIDN.AAP22290.1, PID.q250031
A>Note: sequence extracted from NCBI backbone (NCBI:106452, NCBI:106454)
R.Blatner, F.R., Flunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; M01D:97426617
A:Accession: G64991
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-216 <BIAT>
A:Cross references: GR.AE000019, GR.unn046, NID.q336733, PIDN.AAC75277.1, PID.q1798546
A:Experimental source: strain K 12, substrain MG1655
C:Genetics:
A:Gene: rcsB
A:Map position: 48 min
C:Function:
A:Description: acts as the receiver of effector of the two component regulatory system
A:Pathway: capsule synthesis
C:Superfamily: regulatory protein comA; response regulator homology
C:Keywords: capsule synthesis; DNA binding; phosphoprotein; transcription regulation; tw
F:5-120/Domain: response regulator homology <PRF>
F:56/binding site: phosphate (Asp) (covalent) #status predicted

Query Match 3.5%; Score 7, DB 1, Length 216;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ALAALQK 13
Db 119 ALAALQK 125

Search completed: July 2, 2002, 16:15:39
Job time: 756 sec

A:Reference number: A72451; MUID:994310339
 A:Accession: E72604
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-142 <RAW>
 A:Cross-references: DBJ:AP000061; NID:a5104821; PID:d1044069; PID:g510
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1292

Query Match 3.5% Score 7; DB 2; Length 142;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLSALAA 14
 |||||
 DB 110 LLSALAA 116

RESULT 12

hypothetical protein NMA1494 [imported] Neisseria meningitidis (strain 22491 serogroup
 C:Species: Neisseria meningitidis
 C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: H81840
 R:Parkhill, J.; Achtman, M.; James, K. D.; Bentley, S. D.; Churcher, C.; Klee, S. P.; Mearns
 ; Holtgrave, S.; Jorgensen, K.; Leather, S.; Moule, S.; Muddall, K.; Quail, M. A.; Rajandream
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
 A:Reference number: A41775; MUID:20222556
 A:Accession: H81840
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <PAR>
 A:Cross-references: GAI:162756; GAI:157959; NID:g7380591; PID:CAH84727.1; PID:g738014
 A:Experimental source: serogroup A, strain 22491
 C:Genetics:
 A:Gene: NMA1494
 C:Superfamily: Neisseria meningitidis hypothetical protein NMA1494

Query Match 3.5% Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAA 10
 |||||
 DB 81 LLSALAA 87

RESULT 14

hypothetical protein NMA1284 [imported] Neisseria meningitidis (strain MC58 serogroup
 C:Species: Neisseria meningitidis
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: C81100
 R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Tetlides, A.C.; Nelson, K.E.; Eisen, J.A.
 Hickey, E.K.; Hall, D.H.; Saliberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 et al. Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.D.; Fraser, C.M.; Moxon, F.P.; Pappaji, P.; V
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A1000; MUID:20175755
 A:Accession: C81100
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <REF>
 A:Cross-references: GB:AE002477; GB:AE002098; NID:g722521; PID:AAF41660.1; PID:4722652
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMA1284
 C:Superfamily: Neisseria meningitidis hypothetical protein NMA1494

Query Match 3.5% Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAA 10
 |||||
 DB 81 LLSALAA 87

RESULT 14

conserved hypothetical protein yitE Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: F69839
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, J.; Ber
 C.; Hron, S.; Brouillet, S.; Biuschi, C.V.; Caldwell, R.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Fortati,
 Nature 390, 249-256, 1997
 A:Authors: Fritz, C.; Fujita, N.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Gh
 wood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M
 Koningsstein, G.; Krogh, S.; Kurano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Laube
 A:Authors: Laraviec, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maerli, G.; Mo
 , K.; Ogiwara, A.; Oudoga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Por
 Pivella, C.; Porcia, E.; Porche, B.; Pose, M.; Sadale, Y.; Sato, T.; Sautou, E.; Sch
 A:Authors: Schroeter, K.; Seiflore, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Ser
 amakoshi, A.; Tanaka, T.; Teisflore, F.; Tohno, A.; Totsu, V.; Uchiyama, S.; Vanden
 Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikaw
 A:Authors: Zumbstein, E.; Yoshikawa, H.; Zanchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033
 A:Accession: F69839
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-209 <KUN>
 A:Cross-references: CH:Z99109; CH:AL009126; NID:g263260; PID:CA812946.1; PID:g2644
 A:Experimental source: strain 158
 C:Genetics:
 A:Gene: yitE
 C:Superfamily: conserved hypothetical protein yitE

Query Match 3.5% Score 7; DB 2; Length 209;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 NIKLTSI 56
 |||||
 DB 192 NIKLTSI 198

RESULT 15

colanic acid biosynthesis positive regulator tesB Escherichia coli
 N:Alternate names: regulator of capsule synthesis B component
 C:Species: Escherichia coli
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
 C:Accession: J33668; A44909; G64991
 R:Stout, V.; Gottesman, S.
 J. Bacteriol. 172, 659-669, 1990
 A:Title: tesB and tesC, a two component regulator of capsule synthesis in Escherichia
 A:Reference number: J33668; MUID:90130299
 A:Accession: J33668
 A:Molecule type: DNA
 A:Residues: 1-216 <STO>
 A:Cross-references: GB:M28242; NID:g147524; PID:AAA24504.1; PID:g4457114
 A:Experimental source: strain K12
 R:Gervais, F.G.; Phoenix, P.; Drapau, G.R.
 J. Bacteriol. 174, 3964-3971, 1992
 A:Title: The tesB gene, a positive regulator of colanic acid biosynthesis in Escherich
 A:Reference number: A44909; MUID:92283751

A: Introns: 97/3; 172/3; 837/2; 875/2

Query Match 4.0%; Score 8; DB 2; Length 875;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 130 ENDLEEVG 137
|||||
DB 696 ENDLEEVG 703

RESULT 7

T29390
hypothetical protein K08D10.1 - Caenorhabditis elegans

C: Species: Caenorhabditis elegans
C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C: Accession: T29390

R: (Gelsel), C.; Bradshaw, H.

submitted to the EMBL Data Library, April 1996

A: Description: The sequence of C. elegans cosmid K08D10.

A: Reference number: Z20616

A: Accession: T29390

A: Status: preliminary; Translated from GB/EMBL/DDBJ

A: Molecule type: DNA

A: Residues: 1-935 <GEI>

A: Cross-references: EMBL:U55857; PDB:AAA9635.1, GSPDB:CN00022; ChSP:K08D10.1

A: Experimental source: strain Bristol N2; clone K08D10

C: Genetics:

A: Gene: CESP:K08D10.1

A: Map position: 4

A: Introns: 40/2, 143/3, 218/3; 883/2; 923/3

Query Match 4.0%; Score 8; DB 2; Length 935;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 ENDLEEVG 137
|||||
DB 742 ENDLEEVG 749

RESULT 8

A90691
hypothetical protein PCS0497 [imported] - Escherichia coli (strain O157:H7, substrain R1)

C: Species: Escherichia coli

C: Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C: Accession: A90691

R: Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; Gasawara, T.; Yasunaga, T.; Kubota, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A: Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome

A: Reference number: A90629, NCID:21156231, PMID:11258706

A: Accession: A90691

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-132 <HAY>

A: Cross-references: GB:HA000007; PDB:1AB31920.1; PDB:1AB31920.1; GSPDB:GN00154

A: Experimental source: strain O157:H7, substrain R1MD 0509952

C: Genetics:

A: Gene: ECS0497

C: Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 3.5%; Score 7; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GILSQVI 34
|||||
DB 84 GILSQVI 90

RESULT 9

C64774
ybaW protein - Escherichia coli

C: Species: Escherichia coli

C: Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 24-Nov-1999

C: Accession: C64774

R: Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A: Title: The complete genome sequence of Escherichia coli K-12.

A: Reference number: A64720; M0ID:97426617

A: Accession: C64774

A: Status: nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-132 <BLAT>

A: Cross-references: GB:AF000150; CR:000096; NID:q1786639; PDB:AAK/5546.1; PDB:q17866

A: Experimental source: strain K-12, substrain M5155

C: Genetics:

A: Gene: ybaW

C: Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 3.5%; Score 7; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GILSQVI 34
|||||
DB 84 GILSQVI 90

RESULT 10

R85541
hypothetical protein ybaW [imported] - Escherichia coli (strain O157:H7, substrain ED

C: Species: Escherichia coli

C: Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C: Accession: R85541

R: Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

Miller, L.; Grobbeck, E.J.; Davis, G.; Dimantanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A: Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A: Reference number: A85480; M0ID:21074935; PMID:11206551

A: Accession: R85541

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-132 <STO>

A: Cross-references: GB:AE005174, NID:q12513305; PDB:AAK/54793.1; GSPDB:GN00145; UWA:P

A: Experimental source: strain O157:H7, substrain ED4933

C: Genetics:

A: Gene: ybaW

C: Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 3.5%; Score 7; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GILSQVI 34
|||||
DB 84 GILSQVI 90

RESULT 11

E72603
hypothetical protein APE1292 - Aeropyrum pernix (strain K1)

C: Species: Aeropyrum pernix

C: Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C: Accession: E72603

R: Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jinno, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kadoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A: Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Gene: SUV092

C:Superfamily: cytochrome d complex terminal oxidase chain 1

Query Match 4.0% Score 8; DB 2; Length 467;
 Best Local Similarity 100.0%; Pred. No. 8.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 LTAFLEA 121
 |||||
 Db 102 LTAFLEA 109

RESULT 4

147274

C:Species: pseudomonas aeruginosa

C:Date: 20-Apr-2000 #sequence_revision 20 Apr 2000 #text_change 41 Dec 2000

C:Accession: 147274; GI:4145

R:Gunningham, L.; Pitt, M.; Williams, H.D.

Mol. Microbiol. 24: 579-591, 1997

A:Title: The ctaB genes from Pseudomonas aeruginosa code for a novel cyanide-insensitive

A:Reference number: 74440; MID: 5732424

A:Accession: 147274

A:Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-488 <CON>

A:Cross-references: EMBL:Y10528; NID:32289663; PID:32289664

A:Experimental source: strain PA01, substrain PA0604

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickley, M.J.; Br

adman, S.; Zhao, Y.; Brody, J.C.; Coulter, S.N.; Felger, K.P.; Kas, A.; Lablitz, K.; Lim,

C:Loty, S.; Olson, M.V.

Nature 406: 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A02450; MID: 20437337

A:Accession: GB3155

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-488 <CON>

A:Cross-references: GB:AE004810; GB:AE004091; NID:q9950106; PID:ANG07317.1; GSPDB:GN001

A:Experimental source: strain PA01

A:Gene: ctaA; PA3930

C:Superfamily: cytochrome d complex terminal oxidase chain 1

Query Match 4.0% Score 8; DB 2; Length 488;
 Best Local Similarity 100.0%; Pred. No. 8.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 LTAFLEA 121
 |||||
 Db 102 LTAFLEA 110

RESULT 4

S7384

C:Species: Mycoplasma pneumoniae

C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999

A:Variety: ATCC 29342

R:Hummelreid, R.; Hilbert, H.; Flanagan, H.; Pickl, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24: 4429-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S7384; MID: 97105485

A:Accession: S7384

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-541 <HIM>

A:Cross-references: EMBL:AE000007; GB:0000089; NID:q1674704; PID:AA095/05.1; PID:q167470

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

C:Genetics:

A:Genetic code: SCC3

Query Match 4.0% Score 8; DB 2; Length 541;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AALQKVE 16
 |||||
 Db 50 AALQKVE 57

RESULT 5

A70038

L-lactate permease homolog yviH - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29 Jan 2000

C:Accession: A70038

R:Kunst, F.; Ogasawara, N.; Meszer, L.; Albertini, A.M.; Alloni, G.; Azorido, V.; Her

A:; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, H.; Capuano, V.; Carter, N.M.;

A:; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390: 249-256, 1997

A:Authors: Fowler, D.; Felle, G.; Fujita, M.; Fujita, Y.; Fuma, S.; Galiczi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Iulio, M

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

V.; M.; Ogawa, K.; Ogilwara, A.; Oidega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieget, M.; Rivallid, C.; Roche, F.; Roche, R.; Rose, M.; Sadale, Y.; Sato, T.; Spaul

A:Authors: Schleich, S.; Schropfer, P.; Scollione, P.; Sekiguchi, J.; Sekowska, A.; Se

keuchi, M.; Tanakoshi, A.; Tanaka, T.; Tetsura, P.; Tognoni, A.; Tosiato, V.; Uchiya

i; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.P.; Zumschein, E.; Yoshikawa, H.; Yonehira, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MID: 98044033

A:Accession: A70038

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-563 <KUN>

A:Cross-references: GB:299121; GB:AL009126; NID:32645827; PID:AA15424.1; PID:q46499

A:Experimental source: strain 163

C:Genetics:

A:Gene: yviH

C:Superfamily: L-lactate permease

Query Match 4.0% Score 8; DB 2; Length 563;
 Best Local Similarity 100.0%; Pred. No. 9.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAA 11
 |||||
 Db 15 LLSALAA 22

RESULT 6

I30023

hypoetical protein K08F11.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: I30023

R:Murray, J.; Wohlmann, P.

submitted to the EMBL Data Library, September 1996

A:Description: The sequence of C. elegans cosmid K08F11.

A:Reference number: 220723

A:Accession: I30023

A:Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-875 <EUR>

A:Cross-references: EMBL:070855; PID:AA09160.1; GSPDB:GN00022; GESP:K08F11.2

A:Experimental source: strain Bristol N2; clone K08F11

C:Genetics:

A:Gene: CESP:K08F11.2

A:Map position: 4

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search using sw model

Run on: July 2, 2002, 16:15:38 : Search time 57.09 seconds
(without alignments)
338,307 million cell updates/sec

Title: US-09-603-665-5_COPY_1754_1954

Perfect score: 201

Sequence: 1 EVVLSALAAQKRWETLPHTFYNLADCTAEKLGKLTFL 201

Scoring table:

Gapop 60 0 / Gapext 60 0

Searched: 283138 seqs, 96089334 residues

Word size: 6

Total number of hits satisfying chosen parameters: 1180

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database: PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	Query Match	Length	DR	ID	Description
1	8	4.0	123	2	H75546	hypothetical prote
2	8	4.0	467	2	A75546	probable terminal
3	8	4.0	488	2	T47273	cyanide insensitive
4	8	4.0	541	2	S73383	probable lipoprote
5	8	4.0	563	2	A70038	L-lactate permease
6	8	4.0	875	2	T30023	hypothetical prote
7	8	4.0	935	2	T29390	hypothetical prote
8	7	3.5	132	2	A90691	hypothetical prote
9	7	3.5	132	2	C64774	ybaW protein - hsc
10	7	3.5	132	2	E85541	hypothetical prote
11	7	3.5	132	2	E72603	hypothetical prote
12	7	3.5	146	2	H81840	hypothetical prote
13	7	3.5	146	2	C81100	hypothetical prote
14	7	3.5	209	2	F69839	conserved hypotet
15	7	3.5	216	1	WV0108	colanic acid biosy
16	7	3.5	216	2	D85861	hypothetical prote
17	7	3.5	216	2	A70790	regulator of capsu
18	7	3.5	216	2	R01017	colanic acid biosy
19	7	3.5	217	2	D69271	hypothetical prote
20	7	3.5	217	2	A80149	probable two compo
21	7	3.5	230	2	T12480	hypothetical prote
22	7	3.5	233	2	C71288	probable curies on
23	7	3.5	251	2	H87692	conserved hypotet
24	7	3.5	254	2	E64716	hypothetical prote
25	7	3.5	276	2	T28859	hypothetical prote
26	7	3.5	282	2	A99205	hypothetical prote
27	7	3.5	303	2	T36509	probable molybdopt
28	7	3.5	306	2	G97457	hypothetical prote
29	7	3.5	306	2	A12675	conserved hypotet

ALIGNMENTS

RESULT 1

H75546

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Mar-2000

C:Accession: H75546

R:White, G.; Eisen, J.A.; Heldelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, E.; Attall, T.; Zdzienicka, C.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75390, M010,20036896

A:Accession: H75546

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1 123 <WILL>

A:Cross-references: GB AF001443, GB AF001444, NID q457878, FIDN AAF09801.1; PID:q645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0207

A:Map position: 1

Query Match	4.0%	Score 8:	DB 2:	Length 123:
Best Local Similarity	100.0%	Posed No. 2.4:		
Matches 8:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:

QY 4 LLSALAAAL 11

DB 5 LLSALAAAL 12

RESULT 2

AG0546

probable terminal oxidase chain I [imported] - Salmonella enterica subsp. enterica se

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 03-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001

C:Accession: AG0546

P:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

h, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, E.M.; Dowd, L.; White, N.; Farr

, S.; Mout, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Park, C.; O'Neill, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AG0502; PMID:11677608

A:Accession: AG0546

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-467 <PAR>

A:Cross-references: GB AL513382, FIDN CAC08875.1, PID:AL513382, GSTDB:GN00176

C:Genetics:

09-603-665-5_copy_1867_2067.olig6.rapn

Wed Jul 3 12:05:21 2002

0:

Gaps

0: Indels

0: Mismatches

0: Conservative

Matches

OY 57 AKTDA 62

|||||

DB 90 AKTDA 95

Search completed. July 2, 2002, 16:16:46
Job time: 823 sec


```

1  APPLICANT: Sun, Yungming
2  APPLICANT: Chen, Shu-Yi
3  APPLICANT: Lin, Chueh-Ha
4  APPLICANT: Turner, Leah
5  TITLE OF INVENTION: Compositions and Methods Relating to Linc Specific Genes and Proteins
6  FILE REFERENCE: DEX-0241
7  CURRENT APPLICATION NUMBER: US/10/002 444A
8  CURRENT FILING DATE: 2002-03-15
9  PRIOR APPLICATION NUMBER: US 6/242,998
10 PRIOR FILING DATE: 2000-10-25
11 NUMBER OF SEQ ID NOS: 27
12 SOFTWARE: PatentIn version 4.1
13 SEQ ID NO 241
14 LENGTH: 73
15 TYPE: PRT
16 ORGANISM: Homo sapiens
17 US-10-002-444A-241

Query Match 3.0%; Score 6; DB 6; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 DRLIIF 70
1b 42 DRLIIF 47

RESULT 14
US-10-147-447-730
1 Sequence 730, Application US/1013737
2 GENERAL INFORMATION:
3 APPLICANT: Rosen et al.
4 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
5 FILE REFERENCE: PM2961N
6 CURRENT APPLICATION NUMBER: US/10/137,337
7 CURRENT FILING DATE: 2002-05-03
8 PRIOR APPLICATION NUMBER: 09/758,447
9 PRIOR FILING DATE: 2001-01-11
10 PRIOR APPLICATION NUMBER: 60/179,065
11 PRIOR FILING DATE: 2000-01-31
12 PRIOR APPLICATION NUMBER: 60/180,628
13 PRIOR FILING DATE: 2000-02-04
14 NUMBER OF SEQ ID NOS: 812
15 SOFTWARE: PatentIn Ver. 2.0
16 SEQ ID NO 730
17 LENGTH: 85
18 TYPE: PRT
19 ORGANISM: Homo sapiens
20 US-10-147-447-730

Query Match 3.0%; Score 6; DB 6; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 IAEKIK 82
1b 66 IAEKIK 71

RESULT 14
US-10-106-698-5708
1 Sequence 5708, Application US/10106598
2 GENERAL INFORMATION:
3 APPLICANT: Ruben et al.
4 TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
5 FILE REFERENCE: PA00591
6 CURRENT APPLICATION NUMBER: US/10/106,698
7 CURRENT FILING DATE: 2002-04-27
8 PRIOR APPLICATION NUMBER: PCT/US00/26524
9 PRIOR FILING DATE: 2000-09-28
10 PRIOR APPLICATION NUMBER: US 60/157,137
```

```

1  PRIOR FILING DATE: 1999-09-29
2  PRIOR APPLICATION NUMBER: US 60/163,280
3  PRIOR FILING DATE: 1999-11-03
4  NUMBER OF SEQ ID NOS: 8564
5  SOFTWARE: PatentIn Ver. 3.0
6  SEQ ID NO 5708
7  LENGTH: 92
8  TYPE: PRT
9  ORGANISM: Homo sapiens
10 FEATURE:
11 NAME/KEY: MISC_FEATURE
12 LOCATION: (17)
13 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
14 US-10-106-698-5708

Query Match 3.0%; Score 6; DB 6; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 KTEDAP 63
1b 23 KTEDAP 28

RESULT 15
US-10-023-171-47
1 Sequence 47, Application US/10023171
2 GENERAL INFORMATION:
3 APPLICANT: Weiner, Joel H.
4 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN SECRETION
5 NUMBER OF SEQUENCES: 77
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Medicon & Carroll, LLP
8 STREET: 220 Montgomery Street, Suite 2200
9 CITY: San Francisco
10 STATE: California
11 COUNTRY: United States of America
12 ZIP: 94104
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/10/023,171
20 FILING DATE: 17-Dec-2001
21 CLASSIFICATION: <unknown>
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US/09/085,761
24 FILING DATE: 28-MAY-1998
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Carroll, Peter G.
27 REGISTRATION NUMBER: 32,837
28 REFERENCE/INFORMATION:
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (415) 705-8410
31 TELEFAX: (415) 397-8338
32 INFORMATION FOR SEQ ID NO: 47:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 103 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: Not Relevant
37 TOPOLOGY: unknown
38 MOLECULE TYPE: protein
39 SEQUENCE DESCRIPTION: SEQ ID NO: 47:
40 US-10-023-171-47

Query Match 3.0%; Score 6; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
```

ORGANISM: Saccharomyces cerevisiae
US-60-360-039-1636

Query Match 3.5%; Score 7; DB 7; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLVKPEA 97
|||||
DB 10 HLVKPEA 16

RESULT 8

US-60-360-039-1628
; Sequence 1628, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Harry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1628
; LENGTH: 1612
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-60-360-039-1628

Query Match 3.5%; Score 7; DB 7; Length 1612;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LTFYLA 74
|||||
DB 1265 LTFYLA 1271

RESULT 9

US-60-380-336-395
; Sequence 395, Application US/60380336
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING NOVEL HUMAN G-PROTEIN COUPLED RECEPTORS,
; FILE REFERENCE: D0262 P3F
; CURRENT APPLICATION NUMBER: US/60/380,336
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 395
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-380-336-395

Query Match 3.0%; Score 6; DB 7; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 EEVCKT 26
|||||
DB 1 EEVCKT 6

RESULT 10

US-09-620-393B-550
; Sequence 550, Application US/09620393B
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1068P
; CURRENT APPLICATION NUMBER: US/09/620,393B
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9948
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 550
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..57
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..57
; OTHER INFORMATION: Genes Seq. ID 1376826
US-09-620-393B-550

Query Match 3.0%; Score 6; DB 5; Length 57;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 EEVCKT 26
|||||
DB 19 EEVCKT 24

RESULT 11

US-10-002-344A-240
; Sequence 240, Application US/10002344A
; GENERAL INFORMATION:
; APPLICANT: Recipon, Herve
; APPLICANT: Sun, Yongming
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
; FILE REFERENCE: DEX-0241
; CURRENT APPLICATION NUMBER: US/10/002,344A
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/242,998
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 240
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-002-344A-240

Query Match 3.0%; Score 6; DB 6; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 DRLLTF 70
|||||
DB 42 DRLLTF 47

RESULT 12

US-10-002-344A-241
; Sequence 241, Application US/10002344A
; GENERAL INFORMATION:
; APPLICANT: Recipon, Herve

```
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLKPKPA 97
DB 10 HLKPKPA 16

RESULT 3
US-10-041-018-218
; Sequence 218, Application US/10041018
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041-018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 218
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-218

Query Match 3.5%; Score 7; DB 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLKPKPA 97
DB 10 HLKPKPA 16

RESULT 4
US-10-041-018-264
; Sequence 264, Application US/10041018
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041-018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 264
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-264

Query Match 3.5%; Score 7; DB 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLKPKPA 97
DB 10 HLKPKPA 16

RESULT 5
US-10-041-018-309
; Sequence 309, Application US/10041018
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
```

```
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041-018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 309
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-309

Query Match 3.5%; Score 7; DB 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLKPKPA 97
DB 10 HLKPKPA 16

RESULT 6
US-10-041-018-336
; Sequence 336, Application US/10041018
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041-018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 336
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-336

Query Match 3.5%; Score 7; DB 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 HLKPKPA 97
DB 10 HLKPKPA 16

RESULT 7
US-60-360-039-1836
; Sequence 1836, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1836
; LENGTH: 1045
; TYPE: PRT
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:16:45 : Search time 59.99 seconds
(without alignments)
338.459 Million cell updates/sec

Title: 09-603-665-5_copy_1867_2067
Perfect score: 201
Sequence: 1 LTAFLEALDFRAHSENDL 100 CTAQSFVAMADSLKPLNY 201

Scoring table: OLIGO

Searched: 307805 seqs, 101015595 residues

Gapop 60.0, Gapext 60.0

Word size: 6

Total number of hits satisfying chosen parameters: 254

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database: Pending Patents.MA.New.*

- 1: /cqn2_6/prodata/1/paa/pct.NEW.COMB.pep.*
- 2: /cqn2_6/prodata/1/paa/ms07.NEW.COMB.pep.*
- 3: /cqn2_6/prodata/1/paa/us07.NEW.COMB.pep.*
- 4: /cqn2_6/prodata/1/paa/us08.NEW.COMB.pep.*
- 5: /cqn2_6/prodata/1/paa/us09.NEW.COMB.pep.*
- 6: /cqn2_6/prodata/1/paa/ms10.NEW.COMB.pep.*
- 7: /cqn2_6/prodata/1/paa/ms60.NEW.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	3.5	876	5	US-09-540-209B-7121
2	7	3.5	1045	6	US-10-041-018-199
3	7	3.5	1045	6	US-10-041-018-218
4	7	3.5	1045	6	US-10-041-018-264
5	7	3.5	1045	6	US-10-041-018-309
6	7	3.5	1045	6	US-10-041-018-336
7	7	3.5	1045	7	US-60-360-039-1836
8	7	3.5	1612	7	US-60-360-039-1528
9	6	3.0	6	7	US-60-360-039-395
10	6	3.0	57	7	US-09-540-209B-7121
11	6	3.0	73	6	US-10-002-344A-240
12	6	3.0	74	6	US-10-002-344A-241
13	6	3.0	85	6	US-10-137-337-730
14	6	3.0	92	6	US-10-106-698-5798
15	6	3.0	103	6	US-10-023-171-47
16	6	3.0	103	6	US-10-155-881-21465
17	6	3.0	106	5	US-09-573-655H-460
18	6	3.0	109	5	US-09-882-227-546
19	6	3.0	115	6	US-10-041-018-211
20	6	3.0	115	6	US-10-041-018-351
21	6	3.0	125	5	US-09-667-170A-59
22	6	3.0	144	7	US-60-360-039-8288
23	6	3.0	140	6	US-10-155-881-32557
24	6	3.0	141	6	US-10-117-087-4
25	6	3.0	141	6	US-10-117-087-6
26	6	3.0	160	7	US-60-365-384-255

27	6	3.0	174	6	US-10-155-881-20767
28	6	3.0	202	5	US-09-992-620A-62
29	6	3.0	202	7	US-60-365-384-579
30	6	3.0	211	5	US-09-573-655H-460
31	6	3.0	211	6	US-10-155-881-10291
32	6	3.0	214	7	US-60-360-039-5601
33	6	3.0	220	5	US-09-882-227-410
34	6	3.0	221	6	US-10-160-152-105
35	6	3.0	222	6	US-10-155-881-19909
36	6	3.0	222	6	US-10-155-881-19910
37	6	3.0	222	6	US-10-155-881-32580
38	6	3.0	222	6	US-10-155-881-32609
39	6	3.0	229	6	US-10-155-881-22838
40	6	3.0	231	7	US-60-360-039-5663
41	6	3.0	236	5	US-09-540-209B-7121
42	6	3.0	244	7	US-60-360-039-4608
43	6	3.0	249	7	US-60-360-039-4546
44	6	3.0	251	7	US-60-360-039-17388
45	6	3.0	253	7	US-60-360-039-7304

ALIGNMENTS

RESULT 1
US-09-540-209B-7121
: Sequence 7121, Application US/09540209B
: GENERAL INFORMATION:
: APPLICANT: Gary L. Breton
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
: FILE REFERENCE: 2704/1001-001
: CURRENT APPLICATION NUMBER: US/09/540,209B
: CURRENT FILING DATE: 2000-04-04
: NUMBER OF SEQ ID NOS: 10444
: SEQ ID NO 7121
: LENGTH: 876
: TYPE: PRT
: ORGANISM: H. fragilis
US-09-540-209B-7121

Query Match 3.5%; Score 7; DB 5; Length 876;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 18 NDLEEVG 24
Db 79 NDLEEVG 85

RESULT 2
US-10-041-018-199
: Sequence 199, Application US/10041018
: GENERAL INFORMATION:
: APPLICANT: Matsuda, Seiichi P.T.
: APPLICANT: Hait, Elizabeth A.
: TITLE OF INVENTION: Bacteroides Producing Unicellular Organism
: FILE REFERENCE: P02080051/10025547
: CURRENT APPLICATION NUMBER: US/10/041,018
: CURRENT FILING DATE: 2002-01-07
: PRT-APPLICATION NUMBER: US 09/259880
: PRIOR FILING DATE: 2001-01-05
: NUMBER OF SEQ ID NOS: 413
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 199
: LENGTH: 1045
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-041-018-199

Query Match 3.5%; Score 7; DB 6; Length 1045;



us-09-603-665-5_copy_1754_1954.olig6.rai

Wed Jul 3 09:02:20 2002

Of 46 ASOANIR 52
|
Db 936 ASOANIR 942

Search completed: July 2, 2002, 16:03:59
Job time: 56 sec

TELEFAX: 415-398-3249
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1447 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US 09 656 055 19

Query Match 3.5% Score 7: DB 4: Length 1447;
 Best Local Similarity 100.0% Pred. No. 1.9e+02;
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 46 ASQANIR 52
 |||||
 Db 936 ASQANIR 942

RESULT 14

US 09 954 668 19
 Sequence 19, Application US/0954668
 Patent No. 6172360

GENERAL INFORMATION:

APPLICANT: SCOTT, MATHEW P.
 APPLICANT: GOODRICH, LISA V.
 APPLICANT: JOHNSON, RONALD L.
 TITLE OF INVENTION: Patched Genes and their Use
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley, Hoag & Elliot
 STREET: One Post Office Square
 CITY: Boston
 STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII(TEXT)
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US 954,668

FILING DATE: 20-Oct-1997

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew P.

REGISTRATION NUMBER: 36709

REFERENCE/DOCKET NUMBER: SHV-003.05

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-842-1000

TELEFAX: 617-842-7000

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 1447 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US 09 954 668 19

Query Match 3.5% Score 7: DB 4: Length 1447;
 Best Local Similarity 100.0% Pred. No. 1.9e+02;
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 46 ASQANIR 52
 |||||
 Db 936 ASQANIR 942

RESULT 15

US 09 954 668 19
 Sequence 19, Application US/0954668
 Patent No. 6172360

GENERAL INFORMATION:

APPLICANT: SCOTT, MATHEW P.
 APPLICANT: GOODRICH, LISA V.
 APPLICANT: JOHNSON, RONALD L.
 TITLE OF INVENTION: Patched Genes and their Use
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley, Hoag & Elliot
 STREET: One Post Office Square
 CITY: Boston
 STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII(TEXT)
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US 954,668

FILING DATE: 20-Oct-1997

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew P.

REGISTRATION NUMBER: 36709

REFERENCE/DOCKET NUMBER: SHV-003.05

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-842-1000

TELEFAX: 617-842-7000

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 1447 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US 09 954 668 19

US-09-268-140-5
 Sequence 5, Application US/09268140
 Patent No. 6268176

GENERAL INFORMATION:

APPLICANT: Gemmill, Robert M.
 APPLICANT: Drabkin, Harry A.
 TITLE OF INVENTION: TKCB, A GENE RELATED TO THE HEDGEHOG RECEPTOR, PATCHED
 FILE REFERENCE: 93445-0004
 CURRENT APPLICATION NUMBER: US/09/268,140
 CURRENT FILING DATE: 2000-04-12
 PRIOR APPLICATION NUMBER: US 60/077,723
 PRIOR FILING DATE: 1998-03-12
 NUMBER OF SEQ ID NOS: 46
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 5

LENGTH: 1447

TYPE: PRT

ORGANISM: Homo sapiens

US-09-268-140-5

Query Match 3.5% Score 7: DB 4: Length 1447;
 Best Local Similarity 100.0% Pred. No. 1.9e+02;
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 46 ASQANIR 52
 |||||
 Db 936 ASQANIR 942

RESULT 15

PCT-US95-13233-19

Sequence 19, Application PC/US9513233

GENERAL INFORMATION:

APPLICANT: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY
 TITLE OF INVENTION: Patched Genes and their Use
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: CA

COUNTRY: US

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.40

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/13233

FILING DATE: 06-OCT-1990

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Rowland, Bertram I

REGISTRATION NUMBER: 20015

REFERENCE/DOCKET NUMBER: a60190-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989

TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 1447 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US95-13233-19

Query Match 3.5% Score 7: DB 5: Length 1447;
 Best Local Similarity 100.0% Pred. No. 1.9e+02;
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Query Match 3.5%; Score 7; DB 4; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 ASQANIP 52
| | | | |
Db 922 ASQANIP 928

RESULT 10
PCT-US95-13233-10
; Sequence 10, Application PC/TUS9513233
; GENERAL INFORMATION:
; APPLICANT: THE BOARD OF TRUSTEES OF THE ELAND STANFORD JUNIOR UNIVERSITY
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13233
; FILING DATE: 06-OCT-1990
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-13233-10

Query Match 3.5%; Score 7; DB 5; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 ASQANIP 52
| | | | |
Db 922 ASQANIP 928

RESULT 11
US-08-540-406-19
; Sequence 19, Application US/08540406
; Patent No. 587538
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATTHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco

STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540,406
FILING DATE: 06-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-540-406-19

Query Match 3.5%; Score 7; DB 2; Length 1447;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 ASQANIP 52
| | | | |
Db 936 ASQANIP 942

RESULT 12
US-08-656-055-19
; Sequence 19, Application US/08656055
; Patent No. 6027882
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATTHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/656,055
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/540,406
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989


```

PATENT NO. 5407548
GENERAL INFORMATION:
APPLICANT: SCOTT, MATTHEW P
APPLICANT: GOODRICH, LISA V
APPLICANT: JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540,406
FILING DATE: 06 OCT-1995
CLASSIFICATION: 43E
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1434 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-540-406-10

```

```

Query Match 3.58; Score 7; DB 2; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 46 ASQANIR 52
DB 922 ASQANIR 928

RESULT 8
US-08-656-055-10
Sequence 10, Application US/0865055
Patent No. 6027882
GENERAL INFORMATION:
APPLICANT: SCOTT, MATTHEW P
APPLICANT: GOODRICH, LISA V
APPLICANT: JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/656,055

```

```

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/540,406
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1434 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-656-055-10

```

```

Query Match 3.58; Score 7; DB 3; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 46 ASQANIR 52
DB 922 ASQANIR 928

RESULT 9
US-08-954-668-10
Sequence 10, Application US/08954668
Patent No. 6172200
GENERAL INFORMATION:
APPLICANT: SCOTT, MATTHEW P
APPLICANT: GOODRICH, LISA V
APPLICANT: JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley, Hoag & Elliot
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/954,668
FILING DATE: 20 OCT-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P
REGISTRATION NUMBER: 36709
REFERENCE/DOCKET NUMBER: SHV-004,06
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1434 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-954-668-10

```

DEVELOPMENTAL STAGE: merozoite
US-08-770-035-1

Query Match 3.5%; Score 7; DB 3; Length 315;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAA 10
|||||
DB 56 LLSALAA 62

RESULT 5
US-08-596-366-2
; Sequence 2, Application US/08596366
; Patent No. 5876983
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE,
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER, & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,366
FILING DATE: 29-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-209775
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-209776
FILING DATE: 05-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-784-0 PCT
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 883 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-596-366-2

Query Match 3.5%; Score 7; DB 2; Length 883;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ALOKVVV 16

DB 737 ALOKVVV 743
|||||

RESULT 6
US-08-967-104-2
; Sequence 2, Application US/08967104
; Patent No. 5919694
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE,
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER, & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,104
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,366
FILING DATE: 29-APR-1996
APPLICATION NUMBER: JP 5-209775
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA: JP 5-209776
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-153876
FILING DATE: 05-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-784-0 PCT
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 883 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-104-2

Query Match 3.5%; Score 7; DB 2; Length 883;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ALOKVVV 16
|||||
DB 737 ALOKVVV 743

RESULT 7
US-08-540-406-10
; Sequence 10, Application US/08540406

```

; CORRESPONDENCE ADDRESS
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07110
;
; COMPUTER READABLE FORM
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/729,099
; FILING DATE: 199107 2
; CLASSIFICATION: 424
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Eimeria tenella
; DEVELOPMENTAL STAGE: merozoite
;
; US-08-257-392-1
;
; Query Match 3.5% Score 7: DB 1: Length 315;
; Best Local Similarity 100.0%; Pred. No. 48;
; Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 4 LLSALAA 10
; DB 56 LLSALAA 62
;
; RESULT 4
; US-08-770-035-1
; Sequence 1, Application US/08770035
; Patent No. 6008342
; GENERAL INFORMATION:
; APPLICANT: Binger, Mary-Helen
; APPLICANT: Pasamontes, Luis
; TITLE OF INVENTION: Coccidiosis Vaccines
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/729,099
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Eimeria tenella
;
; US-07-729-099-1
;
; Query Match 4.5% Score 7: DB 1: Length 315;
; Best Local Similarity 100.0%; Pred. No. 48;
; Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 4 LLSALAA 10
; DB 56 LLSALAA 62
;
; RESULT 3
; US-08-257-392-1
; Sequence 1, Application US/08257392
; Patent No. 5688514
; GENERAL INFORMATION:
; APPLICANT: Binger, Mary-Helen
; APPLICANT: Pasamontes, Luis
; TITLE OF INVENTION: Coccidiosis Vaccines
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,392
; FILING DATE: 09 JUN 1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/729,099
; FILING DATE: 12 JUL 1991

```

GenCore version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:03:57; Search time 44.88 seconds
(without alignments)
109.466 Million cell updates/sec

Title: US-09-603-665-5_COPY_1754_1954

Perfect score: 201

Sequence: 1 EVLLSALAALQKVVETLPI..... TFFYNADCTAEKLGKLTFLF 201

Scoring table:

Gapop 60 0, Gapext 60 0

Searched: 231628 seqs, 2442594 residues

Word size: 6

Total number of hits satisfying chosen parameters: 181

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database:

Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCITUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	4.0	829	4	US-09-413-814-105
2	7	3.5	315	1	US-07-720-000-1
3	7	3.5	315	1	US-08-257-392-1
4	7	3.5	315	3	US-08-770-035-1
5	7	3.5	883	2	US-08-546-346-2
6	7	3.5	883	2	US-08-967-104-2
7	7	3.5	1434	2	US-08-540-406-10
8	7	3.5	1434	3	US-08-656-055-10
9	7	3.5	1434	4	US-08-954-668-10
10	7	3.5	1434	5	PCT-US95-13233-10
11	7	3.5	1447	2	US-08-540-406-19
12	7	3.5	1447	3	US-08-656-055-19
13	7	3.5	1447	4	US-08-954-668-14
14	7	3.5	1447	4	US-09-268-140-5
15	7	3.5	1447	5	PCT-US95-13233-19
16	6	3.0	12	3	US-08-578-674-10
17	6	3.0	12	4	US-09-498-346-10
18	6	3.0	29	4	US-04-082-27948-1043
19	6	3.0	29	4	US-09-315-3048-1043
20	6	3.0	31	3	US-08-578-674-8
21	6	3.0	31	4	US-09-498-346-8
22	6	3.0	32	4	US-09-461-697-28
23	6	3.0	34	1	US-08-178-4774-28
24	6	3.0	35	1	US-08-252-9668-9
25	6	3.0	37	3	US-08-578-674-6
26	6	3.0	47	4	US-04-444-444-6
27	6	3.0	43	3	US-08-578-674-7

28 6 3 0 43 4 US-09-498-346-7 Sequence 7, Appli
29 6 3 0 47 3 US-08-578-674-17 Sequence 17, Appli
30 6 3 0 47 4 US-09-498-346-17 Sequence 17, Appli
31 6 3 0 56 3 US-08-578-674-16 Sequence 16, Appli
32 6 3 0 56 4 US-09-498-346-16 Sequence 16, Appli
33 6 3 0 66 3 US-08-578-674-3 Sequence 3, Appli
34 6 3 0 66 3 US-08-578-674-4 Sequence 4, Appli
35 6 3 0 66 3 US-08-578-674-5 Sequence 5, Appli
36 6 3 0 66 4 US-09-498-346-3 Sequence 3, Appli
37 6 3 0 66 4 US-09-498-346-4 Sequence 4, Appli
38 6 3 0 66 4 US-09-498-346-5 Sequence 5, Appli
39 6 3 0 79 2 US-08-839-709-3 Sequence 3, Appli
40 6 3 0 79 2 US-08-839-709-4 Sequence 4, Appli
41 6 3 0 79 2 US-08-204-854-3 Sequence 3, Appli
42 6 3 0 79 2 US-09-204-854-4 Sequence 4, Appli
43 6 3 0 84 3 US-08-578-674-2 Sequence 2, Appli
44 6 3 0 84 4 US-09-498-346-2 Sequence 2, Appli
45 6 3 0 92 4 US-08-965-762-32 Sequence 32, Appli

ALIGNMENTS

RESULT 1

US-09-413-814-105
; Sequence 105, Application US/09413814
; Patent No. 6225064

; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

; APPLICANT: Bristol-Myers Squibb, Co.

; APPLICANT: Reyer, Stefan

; APPLICANT: Bloeker, Helmut

; APPLICANT: Brandt, Petra

; APPLICANT: Cino, Paul M

; APPLICANT: Dougherty, Brian A

; APPLICANT: Goldberg, Steven L

; APPLICANT: Holle, Gerhard

; APPLICANT: Mueller, Joachim

; APPLICANT: Reichenbach, Hans

; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09413814

; EARLIER FILING DATE: 1998-10-07

; EARLIER APPLICATION NUMBER: DE 198 46 493-2

; NUMBER OF SEQ ID NOS: 107

; SOFTWARE: Patent In Ver 2 1

; SEQ ID NO: 105

; LENGTH: 829

; TYPE: PRT

; ORGANISM: Setaugium cellulosum

US-09-413-814-105

Query Match 4.0%; Score 8; DB 4; Length 829;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAAL 11

|||||||

Db 309 LLSALAAL 316

RESULT 2

US-07-720-000-1

; Sequence 1, Application US/07720000

; Patent No. 5404581

; GENERAL INFORMATION:

; APPLICANT: Binger, Mary-Helen

; APPLICANT: Pasamontes, Luis

; TITLE OF INVENTION: Coccidiosis Vaccines

; NUMBER OF SEQUENCES: 15



09-603-665-5_copy_1594_1794.olig6.rapn

wed Jul 3 12:05:17 2002

Db 366 KKEGEE 372

```

RESULT 15
PCT-US02-09671-1124
; Sequence 1124, Application PC/US0209671
; GENERAL INFORMATION:
; APPLICANT: ZYCOS Inc.
; TITLE OF INVENTION: TRANSLATIONAL PROFILING
; FILE REFERENCE: 08191-026W01
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/336,780
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1124
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671-1124

```

Query Match 3.5%; Score 7; DB 1; Length 406;
 Best local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 KKEGEE 69
 Db 364 KKEGEE 370

Search completed: July 2, 2002, 16:16.44
 Job time: 821 sec

```

RESULT 11
PCT-US02-09671-1123
: Sequence 1123, Application PC/TIS0209671
: GENERAL INFORMATION:
: APPLICANT: ZYGEN Inc.
: TITLE OF INVENTION: TRANSLATIONAL PROFILING
: FILE REFERENCE: 08191 026W01
: CURRENT APPLICATION NUMBER: PCT/US02/09671
: CURRENT FILING DATE: 2002-03-28
: PRIOR APPLICATION NUMBER: 60/279,495
: PRIOR FILING DATE: 2001-03-28
: PRIOR APPLICATION NUMBER: 60/292,544
: PRIOR FILING DATE: 2001-05-21
: PRIOR APPLICATION NUMBER: 60/310,801
: PRIOR FILING DATE: 2001-08-08
: PRIOR APPLICATION NUMBER: 60/326,370
: PRIOR FILING DATE: 2001-10-01
: PRIOR APPLICATION NUMBER: 60/336,780
: PRIOR FILING DATE: 2001-12-04
: PRIOR APPLICATION NUMBER: 60/358,985
: PRIOR FILING DATE: 2002-02-20
: NUMBER OF SEQ ID NOS: 2141
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1123
: LENGTH: 490
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US02-09671-1123

```

```

Query Match          3.5%  Score 7;  DB 1;  Length 390;
Best Local Similarity 100.0%;  Pred. No. 1.4e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY 63 KKEGEE 69
| | | | |
DB 464 KKEGEE 370

```

```

RESULT 12
PCT-US02-09671-1128
: Sequence 1128, Application PC/TIS0209671
: GENERAL INFORMATION:
: APPLICANT: ZYGEN Inc.
: TITLE OF INVENTION: TRANSLATIONAL PROFILING
: FILE REFERENCE: 08191 026W01
: CURRENT APPLICATION NUMBER: PCT/US02/09671
: CURRENT FILING DATE: 2002-03-28
: PRIOR APPLICATION NUMBER: 60/279,495
: PRIOR FILING DATE: 2001-03-28
: PRIOR APPLICATION NUMBER: 60/292,544
: PRIOR FILING DATE: 2001-05-21
: PRIOR APPLICATION NUMBER: 60/310,801
: PRIOR FILING DATE: 2001-08-08
: PRIOR APPLICATION NUMBER: 60/326,370
: PRIOR FILING DATE: 2001-10-01
: PRIOR APPLICATION NUMBER: 60/336,780
: PRIOR FILING DATE: 2001-12-04
: PRIOR APPLICATION NUMBER: 60/358,985
: PRIOR FILING DATE: 2002-02-20
: NUMBER OF SEQ ID NOS: 2041
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1128
: LENGTH: 490
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US02-09671-1128

```

```

Query Match          3.5%  Score 7;  DB 1;  Length 390;
Best Local Similarity 100.0%;  Pred. No. 1.4e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY 63 KKEGEE 69
| | | | |
DB 364 KKEGEE 370

RESULT 13
PCT-US02-09671-1125
: Sequence 1125, Application PC/TIS0209671
: GENERAL INFORMATION:
: APPLICANT: ZYGEN Inc.
: TITLE OF INVENTION: TRANSLATIONAL PROFILING
: FILE REFERENCE: 08191 026W01
: CURRENT APPLICATION NUMBER: PCT/US02/09671
: CURRENT FILING DATE: 2002-03-28
: PRIOR APPLICATION NUMBER: 60/279,495
: PRIOR FILING DATE: 2001-03-28
: PRIOR APPLICATION NUMBER: 60/292,544
: PRIOR FILING DATE: 2001-05-21
: PRIOR APPLICATION NUMBER: 60/310,801
: PRIOR FILING DATE: 2001-08-08
: PRIOR APPLICATION NUMBER: 60/326,370
: PRIOR FILING DATE: 2001-10-01
: PRIOR APPLICATION NUMBER: 60/336,780
: PRIOR FILING DATE: 2001-12-04
: PRIOR APPLICATION NUMBER: 60/358,985
: PRIOR FILING DATE: 2002-02-20
: NUMBER OF SEQ ID NOS: 2041
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1125
: LENGTH: 392
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US02-09671-1125

```

```

Query Match          3.5%  Score 7;  DB 1;  Length 392;
Best Local Similarity 100.0%;  Pred. No. 1.5e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY 63 KKEGEE 69
| | | | |
DB 366 KKEGEE 372

```

```

RESULT 14
US-10-102-806-510
: Sequence 510, Application US/19182806
: GENERAL INFORMATION:
: APPLICANT: ROSEN et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PAL03P1C1
: CURRENT APPLICATION NUMBER: US/10/102,806
: CURRENT FILING DATE: 2002-03-22
: PRIOR APPLICATION NUMBER: 09/925,298
: PRIOR FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05881
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 846
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 510
: LENGTH: 392
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-102-806-510

```

```

Query Match          3.5%  Score 7;  DB 6;  Length 392;
Best Local Similarity 100.0%;  Pred. No. 1.5e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY 63 KKEGEE 69

```

```

; Sequence 10675, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; TITLE OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P US/09/935,625
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 10675
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..138
; OTHER INFORMATION: Ceres Seq. ID no. 1017008
US-09-935-625-10675

```

```

Query Match          3.5%: Score 7; DB 5; Length 138;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 54 DLLAIQV 60
DB 132 DLLAIQV 138

```

```

RESULT 8
; Sequence 6605, Application US/09620393B
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1068P
; CURRENT APPLICATION NUMBER: US/09/620,393B
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9948
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6605
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..158
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..158
; OTHER INFORMATION: Ceres Seq. ID 1393051
US-09-620-393B-6605

```

```

Query Match          3.5%: Score 7; DB 5; Length 158;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 165 LSALAAL 171
DB 86 LSALAAL 92

```

```

RESULT 9
; Sequence 3230, Application US/60460039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.

```

```

; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/460,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2220
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-60-460-039-2220

```

```

Query Match          3.5%: Score 7; DB 7; Length 316;
Best Local Similarity 100.0%; Pred. No. 1,26+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 TPIPVIR 12
DB 31 TPIPVIR 37

```

```

RESULT 10
; PCT-US02-13329-8
; Sequence 8, Application PC/TUS0213329
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: THORNTON, Michael
; APPLICANT: WALIA, Nalinder K.
; APPLICANT: GRIETZEN, Kimberly J.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: YAO, Monique G.
; APPLICANT: JIN, Pei
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: HOROWSKY, Mark L.
; APPLICANT: GRAULI, Richard C.
; APPLICANT: YANG, Junming
; APPLICANT: DING, Li
; APPLICANT: FU, Glenn K.
; TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: PI-0415 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/13329
; CURRENT FILING DATE: 2002-04-25
; PCT APPLICATION NUMBER: 60/287,151, 60/260,516, 60/231,217, 60/214,972;
; 60/429,217; 60/434,718; 60/443,903
; PRIOR FILING DATE: 2001-04-27; 2001-05-11; 2001-05-15; 2001-08-24;
; 2001-10-12; 2001-10-19; 2001-11-02
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7472061CD1
PCT-US02-13329-8

```

```

Query Match          3.5%: Score 7; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 1,26+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 162 VYLSAL 168
DB 42 VYLSAL 48

```


OTHER INFORMATION: Ceres Seq. ID no. 3086968
US-09-945-625-9918

Query Match 4.0% Score 8; DB 5; Length 603;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKEEEE 69
| | | | |
DB 2 KKKKEEEE 9

RESULT 3

US-09-945-625-9918
Sequence 9918, Application US/09945625

GENERAL INFORMATION:

APPLICANT: N. ALEXANDROV et al.

TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE

TITLE OF INVENTION: MODULATING VARIOUS RESPONSES

FILE REFERENCE: 2750-1481P

CURRENT APPLICATION NUMBER: US/09/945-625

CURRENT FILING DATE: 2001-08-24

NUMBER OF SEQ ID NOS: 33136

SEQ ID NO 9918

LENGTH: 606

TYPE: PRT

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: peptide

LOCATION: 1..606

OTHER INFORMATION: Ceres Seq. ID no. 3086967

US-09-945-625-9918

Query Match

Best Local Similarity 4.0%; Score 8; DB 5; Length 606;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKEEEE 69
| | | | |
DB 5 KKKKEEEE 12

RESULT 4

US-60-482-898-332

Sequence 332, Application US/60482898

GENERAL INFORMATION:

APPLICANT: Hudson, Keith

APPLICANT: et al.

TITLE OF INVENTION: Plant Receptors and Ligands

FILE REFERENCE: 1066P

CURRENT APPLICATION NUMBER: US/60/482-898

CURRENT FILING DATE: 2002-05-22

NUMBER OF SEQ ID NOS: 1344

SOFTWARE: FastSeq for Windows Version 4.1

SEQ ID NO 332

LENGTH: 664

TYPE: PRT

ORGANISM: Arabidopsis thaliana

US-60-482-898-332

Query Match

Best Local Similarity 4.0%; Score 8; DB 7; Length 664;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKEEEE 69
| | | | |
DB 567 KKKKEEEE 574

RESULT 5

US-09-620-3938-6607
Sequence 6607, Application US/096203938

GENERAL INFORMATION:

APPLICANT: ALEXANDROV, Nikolai et al.

TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID

FILE REFERENCE: 2750-1068P

CURRENT APPLICATION NUMBER: US/09/620-3938

CURRENT FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 9948

SOFTWARE: PatentIn version 3.0

SEQ ID NO 6607

LENGTH: 105

TYPE: PRT

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..105

OTHER INFORMATION: Xaa is any amino acid

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..105

OTHER INFORMATION: Ceres Seq. ID 1393053

US-09-620-3938-6607

Query Match

Best Local Similarity 3.5%; Score 7; DB 5; Length 105;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 LSALAAL 171
| | | | |
DB 33 LSALAAL 39

RESULT 6

US-09-620-3938-6606

Sequence 6606, Application US/096203938

GENERAL INFORMATION:

APPLICANT: ALEXANDROV, Nikolai et al.

TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID

FILE REFERENCE: 2750-1068P

CURRENT APPLICATION NUMBER: US/09/620-3938

CURRENT FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 9948

SOFTWARE: PatentIn version 3.0

SEQ ID NO 6606

LENGTH: 121

TYPE: PRT

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..121

OTHER INFORMATION: Xaa is any amino acid

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..121

OTHER INFORMATION: Ceres Seq. ID 1393052

US-09-620-3938-6606

Query Match

Best Local Similarity 3.5%; Score 7; DB 5; Length 121;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 LSALAAL 171
| | | | |
DB 49 LSALAAL 55

RESULT 7

US-09-945-625-10675

GenCore version 4.5
Copyright (c) 1993 2000 Compugen Ltd

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:03:03 : Search time 59.99 seconds
(without alignments)
338,459 million cell updates/sec

Title: US-09-603-665-5_COPY_1594_1794

Perfect score: 231

Sequence: 1 LLPTETPVPRLV:NPLP ISPVLEILSQVHLEKITS 201

Scoring table: Gapped

Gapop 60.0, Gapext 60.0

Searched: 307805 seqs, 101015545 residues

Word size: 6

Total number of hits satisfying chosen parameters: 1858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database: Pending Patents_AA_New*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	4	0	543	Sequence 20635, A
2	8	4	0	604	Sequence 9919, Ap
3	8	4	0	606	Sequence 9918, Ap
4	8	4	0	664	Sequence 332, App
5	7	3.5	105	7 US-60-382-898-332	Sequence 6607, Ap
6	7	3.5	121	5 US-09-620-393B-6606	Sequence 6606, Ap
7	7	3.5	138	5 US-09-935-625-10675	Sequence 10675, A
8	7	3.5	158	5 US-09-620-393B-6605	Sequence 6605, Ap
9	7	3.5	316	7 US-60-360-039-4220	Sequence 4220, Ap
10	7	3.5	323	1 PCT-US02-13329.8	Sequence 8, Appli
11	7	3.5	390	1 PCT-US02-09671-1123	Sequence 1123, Ap
12	7	3.5	390	1 PCT-US02-09671-1128	Sequence 1128, Ap
13	7	3.5	392	1 PCT-US02-09671-1125	Sequence 1125, Ap
14	7	3.5	392	6 US-10-102-806-510	Sequence 510, App
15	7	3.5	406	1 PCT-US02-09671-1124	Sequence 1124, Ap
16	7	3.5	406	1 PCT-US02-09671-1126	Sequence 1126, Ap
17	7	3.5	406	1 PCT-US02-09671-1127	Sequence 1127, Ap
18	7	3.5	409	7 US-60-360-039-429	Sequence 929, App
19	7	3.5	452	5 US-09-573-655B-2123	Sequence 2123, Ap
20	7	3.5	452	5 US-09-540-209B-9330	Sequence 9330, Ap
21	7	3.5	473	5 US-09-573-655B-239	Sequence 239, App
22	7	3.5	473	5 US-09-573-655B-786	Sequence 786, App
23	7	3.5	579	5 US-10-041-038-261	Sequence 261, App
24	7	3.5	688	7 US-60-360-039-6112	Sequence 6112, Ap
25	7	3.5	727	1 PCT-US02-09944.649	Sequence 649, Appl
26	7	3.5	813	7 US-60-360-039-12614	Sequence 12614, A

27 7 3.5 818 7 US-60-360-039-5568
28 7 3.5 850 1 PCT-US02-13142-3534
29 7 3.5 850 6 US-10-128-714-3534
30 7 3.5 872 7 US-60-460-039-3832
31 7 3.5 883 1 PCT-US02-13142-8534
32 7 3.5 883 6 US-10-128-714-8534
33 7 3.5 883 7 US-60-360-039-23646
34 7 3.5 1133 7 US-60-360-039-4112
35 7 3.5 1161 7 US-60-460-039-3660
36 6 3.0 11 6 US-10-033-741-53
37 6 3.0 11 6 US-10-033-662-48
38 6 3.0 20 6 US-10-105-299-5116
39 6 3.0 42 5 US-09-620-393B-6608
40 6 3.0 55 6 US-10-155-881-9573
41 6 3.0 61 6 US-10-002-344A-231
42 6 3.0 62 5 US-09-540-209B-786
43 6 3.0 64 6 US-10-155-881-9610
44 6 3.0 67 6 US-10-157-031-156
45 6 3.0 79 6 US-10-155-881-12614

ALIGNMENTS

RESULT 1
US-60-360-039-20635
Sequence 20635, Application US/60360039
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Chen, Xianfeng
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)A
CURRENT APPLICATION NUMBER: US/60/460,039
CURRENT FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 20635
LENGTH: 543
TYPE: PRT
ORGANISM: Rhodospseudomonas palustris
US-60-360-039-20635

Query Match 4.0%; Score 8; DB 7; Length 543;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8, Conservative 0, Mismatches 0, Indels 0; Gaps 0;

QY 164 LLSALAAAL 171
DB 177 LLSALAAAL 184

RESULT 3

US-09-935-625-9919
Sequence 9919, Application US/09935625
GENERAL INFORMATION:
APPLICANT: N. ALEXANDROV et al.
TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAFA
FILE REFERENCE: 2750-1481P
CURRENT APPLICATION NUMBER: US/09/935,625
CURRENT FILING DATE: 2001-08-24
NUMBER OF SEQ ID NOS: 33136
SEQ ID NO 9919
LENGTH: 603
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: peptide
LOCATION: 1-603

10 103 LTAFLEA 110

RESULT 14

US 09 543 681A-6481
Sequence 6481, Application US/09/543681A

GENERAL INFORMATION:

APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEINS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543.681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999 04-09
NUMBER OF SEQ ID NOS: 8444

SEQ ID NO 6481

LENGTH: 457

TYPE: PRT

ORGANISM: Proteus mirabilis

US 09 543 681A 6481

Query Match

Best Local Similarity 4.0%; Score 8; DB 19; Length 457;

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 LTAFLEA 8

DB 110 LTAFLEA 117

RESULT 14

US 09 897-516-7795

Sequence 7795, Application US/09/897516

GENERAL INFORMATION:

APPLICANT: Corbin, David R.
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Hueslod, Joseph E.
APPLICANT: Krasomil-Osterfeld, Karina C.
APPLICANT: Malvar, Thomas M.
APPLICANT: Slater, Steven C.
APPLICANT: Spiridonov, Sergei

TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof

FILE REFERENCE: 38-21(51847)B

CURRENT APPLICATION NUMBER: US/09/897,516

CURRENT FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 60/215, 161

PRIOR FILING DATE: 2000 06-30

NUMBER OF SEQ ID NOS: 8409

SEQ ID NO 7795

LENGTH: 460

TYPE: PRT

ORGANISM: Xenorhabdus sp.

US 09 897-516-7795

Query Match

Best Local Similarity 4.0%; Score 8; DB 22; Length 460;

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 LTAFLEA 8

DB 103 LTAFLEA 110

RESULT 16

US 60 215 161 7795

Sequence 7795, Application US/60/215161

GENERAL INFORMATION:

APPLICANT: Corbin, David R.

APPLICANT: Goldman, Barry S.

APPLICANT: Huesing, Joseph E.
APPLICANT: Krasomil-Osterfeld, Karina C.
APPLICANT: Malvar, Thomas M.
APPLICANT: Slater, Steven C.
APPLICANT: Spiridonov, Sergei
APPLICANT: Hinkle, Gregory J.
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
FILE REFERENCE: 38-21(51847)A
CURRENT APPLICATION NUMBER: US/60/215,161
CURRENT FILING DATE: 2000-06-30
NUMBER OF SEQ ID NOS: 8409
SEQ ID NO 7795
LENGTH: 460
TYPE: PRT
ORGANISM: Xenorhabdus sp.

US-60-215-161-7795

Query Match

Best Local Similarity 4.0%; Score 8; DB 26; Length 460;

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 LTAFLEA 8

DB 103 LTAFLEA 110

Search completed: July 2, 2002, 16:14:37

Job time: 694 sec

SEQ ID NO 6
LENGTH: 220
TYPE: PRT
ORGANISM: Chitinophaga pinensis
US-09-823-823-6

Query Match 4.0%, Score 8, DB 22, Length 220;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SENDPEKC 122
Db 130 SENDPEKC 137
|||||

RESULT 9
US-09-823-829-6
; Sequence 6, Application US/09823829
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Satoshi
; APPLICANT: Nakamura, Shoko
; APPLICANT: Suzuki, Makoto
; APPLICANT: Kasai, Hiroaki
; APPLICANT: Hamada, Tohru
; TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS
; FILE REFERENCE: 12817-004001
; CURRENT APPLICATION NUMBER: US/09/823,829
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/208,688
; PRIOR FILING DATE: 1998-12-10
; PRIOR APPLICATION NUMBER: JP 97/343316
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 6
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Chitinophaga pinensis
US-09-823-829-6

Query Match 4.0%, Score 8, DB 22, Length 220;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SENDPEKC 122
Db 130 SENDPEKC 137
|||||

RESULT 10
US-09-489-039A-7912
; Sequence 7912, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Barton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709, 2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7912
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7912

Query Match 4.0%, Score 8, DB 18, Length 411;

Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFFLEA 8
Db 46 LTAFFLEA 53
|||||

RESULT 11
US-09-897-515-5029
; Sequence 5029, Application US/09897515
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 5029
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-5029

Query Match 4.0%, Score 8, DB 22, Length 445;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFFLEA 8
Db 103 LTAFFLEA 110
|||||

RESULT 12
US-60-215-161-5029
; Sequence 5029, Application US/60215161
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)A
; CURRENT APPLICATION NUMBER: US/60/215,161
; CURRENT FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 5029
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-60-215-161-5029

Query Match 4.0%, Score 8, DB 26, Length 445;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFFLEA 8
|||||

PRIOR FILING DATE: 2000-02-24
 PRIOR APPLICATION NUMBER: US 09/577,409
 PRIOR FILING DATE: 2000-05-18
 PRIOR APPLICATION NUMBER: PCT/US01/34,427
 PRIOR FILING DATE: 2001-02-26
 PRIOR APPLICATION NUMBER: US 09/519,705
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: US 09/574,454
 PRIOR FILING DATE: 2000-05-19
 PRIOR APPLICATION NUMBER: PCT/US01/04941
 PRIOR FILING DATE: 2001-04-05
 PRIOR APPLICATION NUMBER: US 09/540,217
 PRIOR FILING DATE: 2000-03-31
 PRIOR APPLICATION NUMBER: US 09/549,167
 PRIOR FILING DATE: 2000-08-23
 PRIOR APPLICATION NUMBER: PCT/US01/08631
 PRIOR FILING DATE: 2001-04-30
 PRIOR APPLICATION NUMBER: US 09/552,929
 PRIOR FILING DATE: 2000-04-18
 PRIOR APPLICATION NUMBER: US 09/770,160
 PRIOR FILING DATE: 2001-01-26
 PRIOR APPLICATION NUMBER: PCT/US01/08656
 PRIOR FILING DATE: 2001-04-18
 NUMBER OF SEQ ID NOS: 412
 SOFTWARE: P2_Fligenes Version 6.0
 SEQ ID NO 167
 LENGTH: 1149
 TYPE: PRT
 ORGANISM: Homo Sapiens
 US 60 349 454 167

Query Match 49.8% Score 100; DB 26; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 1,66-94;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LTAFLEALDFAQHSNDLEEVGKTENCIIICLVAMVVKLSEVTPPLFFKLFDAWAKE 60
 DB 872 LTAFLEALDFAQHSNDLEEVGKTENCIIICLVAMVVKLSEVTPPLFFKLFDAWAKE 931
 QY 61 DAPKRDLLTFYNLADCIATKLGKLFILFAGHLVKPFADTL 100
 DB 942 DAPKRDLLTFYNLADCIATKLGKLFILFAGHLVKPFADTL 971

RESULT 7
 US 60-141-424-4
 Sequence 3, Application US/60141423
 GENERAL INFORMATION:
 APPLICANT: Barry, Caroline
 APPLICANT: Boudegeret, Lydie
 APPLICANT: Chumakov, Ilya
 TITLE OF INVENTION: A NVEL DAP28 GENE AND PROTEIN
 FILE REFERENCE: GENSET 063PPE
 CURRENT APPLICATION NUMBER: US/60/141-424
 CURRENT FILING DATE: 1999-06-25
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: Patent.pm
 SEQ ID NO 3
 LENGTH: 2144
 TYPE: PRT
 ORGANISM: homo sapiens
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 91..111
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 212..242
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 700..720

OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 861..881
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 884..904
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 911..931
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 947..957
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 1297..1317
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 1410..1430
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 1725..1745
 OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: TRANSMEM
 LOCATION: 1987..2007
 OTHER INFORMATION: potential membrane spanning segment, TopPred Program
 FEATURE:
 NAME/KEY: SITE
 LOCATION: 1750..1771
 OTHER INFORMATION: potential leucine zipper pattern, BIA Program
 US 60-141-323-3

Query Match 49.8% Score 100; DB 26; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 2,90-94;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFLEALDFAQHSNDLEEVGKTENCIIICLVAMVVKLSEVTPPLFFKLFDAWAKE 60
 DB 1867 LTAFLEALDFAQHSNDLEEVGKTENCIIICLVAMVVKLSEVTPPLFFKLFDAWAKE 1926
 QY 61 DAPKRDLLTFYNLADCIATKLGKLFILFAGHLVKPFADTL 100
 DB 1927 DAPKRDLLTFYNLADCIATKLGKLFILFAGHLVKPFADTL 1966

RESULT 8
 US-09-823-823-6
 Sequence 6, Application US/09823823
 GENERAL INFORMATION:
 APPLICANT: Yamamoto, Satoshi
 APPLICANT: Kasai, Hiroaki
 APPLICANT: Nakamura, Shoko
 APPLICANT: Suzuki, Makoto
 APPLICANT: Hamada, Tohru
 TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS USING
 FILE REFERENCE: 12817-004001
 CURRENT APPLICATION NUMBER: US/09/823,823
 CURRENT FILING DATE: 2001-03-30
 PRIOR APPLICATION NUMBER: US 09/208,688
 PRIOR FILING DATE: 1998-12-10
 PRIOR APPLICATION NUMBER: JP 97/343316
 PRIOR FILING DATE: 1997-12-12
 NUMBER OF SEQ ID NOS: 80
 SOFTWARE: Patent in version 2.0

```

QY 61 DAPKDRLLTFYNIADCIAEKLGKGLFTLPFAGHLVKKPADTL 100
Db 298 DAPKDRLLTFYNIADCIAEKLGKGLFTLPFAGHLVKKPADTL 337

RESULT 4
US-60-042-611-52
; Sequence 52, Application US/60042611
; GENERAL INFORMATION:
; APPLICANT: Bowcock, Anne M.
; APPLICANT: Baer, Richard
; TITLE OF INVENTION: Compositions and Methods Comprising
; TITLE OF INVENTION: BARD1 and Other BRCA1 Binding Proteins
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/042,611
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: P-41,071
; REFERENCE/DOCKET NUMBER: HTSD-499P22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-60-042-611-52

Query Match 49.8%; Score 100; DB 26; Length 515;
Best Local Similarity 100.0%; Pred. No. 7.8e-95;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFLEALDFRAQSENDEEYGVKTENCIDCLVAMVVKLSKVTFRPLFFKLFQWAKTE 60
Db 238 LTAFLEALDFRAQSENDEEYGVKTENCIDCLVAMVVKLSKVTFRPLFFKLFQWAKTE 297

QY 61 DAPKDRLLTFYNIADCIAEKLGKGLFTLPFAGHLVKKPADTL 100
Db 298 DAPKDRLLTFYNIADCIAEKLGKGLFTLPFAGHLVKKPADTL 337

RESULT 5
US-60-042-985-52
; Sequence 52, Application US/60042985
; GENERAL INFORMATION:
; APPLICANT: Bowcock, Anne M.
; APPLICANT: Baer, Richard
; TITLE OF INVENTION: Compositions and Methods Comprising
; TITLE OF INVENTION: BARD1 and Other BRCA1 Binding Proteins
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/042,985
; FILING DATE: Concurrently herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: P-41,071
; REFERENCE/DOCKET NUMBER: HTSD-499P23
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-60-042-985-52

Query Match 49.8%; Score 100; DB 26; Length 515;
Best Local Similarity 100.0%; Pred. No. 7.8e-95;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFLEALDFRAQSENDEEYGVKTENCIDCLVAMVVKLSKVTFRPLFFKLFQWAKTE 60
Db 238 LTAFLEALDFRAQSENDEEYGVKTENCIDCLVAMVVKLSKVTFRPLFFKLFQWAKTE 297

QY 61 DAPKDRLLTFYNIADCIAEKLGKGLFTLPFAGHLVKKPADTL 100
Db 298 DAPKDRLLTFYNIADCIAEKLGKGLFTLPFAGHLVKKPADTL 337

RESULT 6
US-60-339-453-167
; Sequence 167, Application US/60339453
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Weng, Gechi
; APPLICANT: Boyle, Bryan J
; APPLICANT: Dmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 812
; CURRENT APPLICATION NUMBER: US/60/339,453
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/441,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126

```

```

1 LOCATION: 1967
2 OTHER INFORMATION: Xaa-Asp or Asn
3 NAME/KEY: VARIANT
4 LOCATION: 2017
5 OTHER INFORMATION: Xaa-Gly or Glu
6 Z-09 603-665-5

```

Query Match	99.08;	Score	199;	DB	20;	Length	2144;
Best Local Similarity	00.08;	Pred. No.	28-197;				
Matches 201;	Conservat	ve	0;	Mismatches	0;	Indels	0;
						Caps	0;

1 1TAPLEALDJEKACHSINDLEEVCKTENC11007LVAMVVKLSEVTFPRPEKIFDWAKE 60
1867 11A1FTEALDJEKACHSINDLEEVCKTENC11007LVAMVVKLSEVTFPRPEKIFDWAKE 1926

Q7 61 DAPKRLITPYNLADCI AEKI KGLFTL FAGHLVKPFADTL XOVNI SKTDPAFFUSNDPE 120
 |||||
b 1927 GAKPKRLITPYNLADCI AEKI KGLFTL FAGHLVKPFADTL XOVNI SKTDPAFFUSNDPE 1886

121 KCTLLQFTNCLYKTFLEFQHFISNKRAXALMPLVIOLENLGGEEKFQERVTKHLI 180
 |||||
 1987 KCTLLQFTNCLYKTFLEFQHFISNKRAXALMPLVIOLENLGGEEKFQERVTKHLI 2045

181 PCTAGFSVANAIGSLWKP.LNY 201
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247 PCTAGFSVANAIGSLWKP.LNY 2067
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688

RESULT	2	4
US	60	176
	880	

; APPLICANT: Harry, Caroline
 ; APPLICANT: Bouqueloiret, Lydie
 ; GENERAL INFORMATION:
 ; Acquisition 4: Application CS/9017-9660

; REFERENCE: CHOMAROV, ILYA
 ; TITLE OF INVENTION: A NOVEL BAP28 GENE AND PROTEIN.
 ; FILE REFERENCE: 67.US2.FRO
 ; CURRENT APPLICATION NUMBER: US/60/176,880

```

; CURRENT FILING DATE: 2000-01-18
; EARLIER APPLICATION NUMBER: US 60/141,323
; EARLIER FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 12

```

```

; SEQ ID NO 4
;     LENGTH: 2144
;     TYPE: PRO

```

1	NAME:	Xaa-Ser
2	FEATURE:	VARIANT
3	NAME/KEY:	VARIANT
4	LOCATION:	1694
5	OTHER INFORMATION:	Xaa-Ser or Asn
6	FEATURE:	
7	NAME/KEY:	VARIANT
8	LOCATION:	1654
9	OTHER INFORMATION:	Xaa-Ala or Val
10	FEATURE:	
11	NAME/KEY:	VARIANT
12	LOCATION:	1967
13	OTHER INFORMATION:	Xaa-Asp or Asn
14	FEATURE:	
15	NAME/KEY:	VARIANT
16	LOCATION:	2017
17	OTHER INFORMATION:	Xaa-Tyr or Glu
18	FEATURE:	
19	NAME/KEY:	VARIANT
20	LOCATION:	176-880-4

Query Match 93.0%; Score 199; DB 26; Length 2144;
 Best local Similarity 130.0%; Pred. No. 26-197;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LTAPFLFAIDERRAHSNDMEFVAKTENCTITP'VAMVVI SEVTFPP'PFKLEFVAKTE 60

1867	DB	LTAFTEALDFPAGHSENEUFEV:KTEI ⁰ IILU ⁰ LVAMVVKSEVITFPLEFFKLEFWAKTE	1926
61	QY	DAPKORU ⁰ TFYNAUCIAEKIKGLFTTFAGHVKPFAU ⁰ IXONI:SKTDEAFUSINDOE	120
1927	DB	DAPKORU ⁰ TFYNAUCIAEKIKGLFTTFAGHVKPFAU ⁰ IXONISKI:DEAFUSINDOE	1986
121	QY	KCCLLIQF ⁰ ILNLYK ⁰ FTFDT ⁰ HFISKEPAXALMPLVQLENP ⁰ LGEFKEFVKHLL	180
1987	DB	KCCLLIQF ⁰ ILNCLYK ⁰ FTFDT ⁰ HFISKEPAXALMPLVQLENP ⁰ LGEFKEFVKHLL	2046
181	QY	PCIAQFSVAMADSLWKPNY	201
2047	DB	PCIAQFSVAMADSLWKPNY	2067

```

1  RESULT 3
2  US-08-936-487-52
3  : Sequence 52, Application US/01936487
4  : GENERAL INFORMATION:
5  : APPLICANT: Bowcock, Anne M.
6  : APPLICANT: Baer, Richard
7  : TITLE OF INVENTION: Compositions and Methods Comprising
8  : TITLE OF INVENTION: HARDI and Other BRCAL Binding Proteins
9  : NUMBER OF SEQUENCES: 130
10 : CORRESPONDENCE ADDRESS:
11 : ADDRESSEE: Arnold, White & Durkee
12 : STREET: P.O. Box 4433
13 : CITY: Houston
14 : STATE: Texas
15 : COUNTRY: USA
16 : ZIP: 77210
17 : COMPUTER READABLE FORM:
18 : MEDIUM TYPE: Floppy disk
19 : COMPUTER: IBM PC compatible
20 : OPERATING SYSTEM: PC-DOS/MS-DOS
21 : SOFTWARE: Patent In Release #1.0, Version #1.30
22 : CURRENT APPLICATION DATA:
23 : APPLICATION NUMBER: US/08/936.487
24 : FILING DATE: Concurrent y Herewith
25 : CLASSIFICATION: 435
26 : PRIOR APPLICATION DATA:
27 : APPLICATION NUMBER: US 60/025,296
28 : FILING DATE: 20 SEP-1996
29 : PRIOR APPLICATION DATA:
30 : APPLICATION NUMBER: US 60/042,611
31 : FILING DATE: 03-APR-1997
32 : PRIOR APPLICATION DATA:
33 : APPLICATION NUMBER: US 60/042,985
34 : FILING DATE: 04-APR-1997
35 : ATTORNEY/AGENT INFORMATION:
36 : NAME: Hibler, David W.
37 : REGISTRATION NUMBER: P-41-071
38 : REFERENCE/DOCKET NUMBER: UIUSD:499
39 : TELECOMMUNICATION INFORMATION:
40 : TELEPHONE: 512/418-3000
41 : TELEFAX: 512/474-7577
42 : INFORMATION FOR SEQ ID NO: 52:
43 : SEQUENCE CHARACTERISTICS:
44 : LENGTH: 515 amino acids
45 : TYPE: amino acid
46 : STRANDEDNESS:
47 : TOPOLOGY: linear
48 : US-08-936-487-52

```

Query Match 49.8%; Score 100; DB 13; Length 515;
Best Local Similarity 100.0%; Pred. No. 7.8e-95;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gap-S 0;

Qy 1 LTAFLEAFJEPRAQHSNDLFVKGTFNC11DC1VAMVVKLSEVTERPLEFKLFIWAKTE 60

.....

Genome version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:14:35, Search time 505.4 seconds
(without alignments)
146,012 million cell updates/sec

Title: 09-603-665-5_copy_1867_2067

Perfect score: 201

Sequence: 1 LTAFLEALDFRAGISENC CTAQFSVAMADSLWKIPNY 201

Scoring table: OLI60
Gapop 60.0, Gapext 60.0

Searched: 3502263 seqs, 351980561 residues

Word size: 6

Total number of hits satisfying chosen parameters: 1663

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database: Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US105_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US108_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US109_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US110_COMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199	99.0	2144	20	US-09-603-665-5
2	199	99.0	2144	26	US-60-176-880-4
3	100	49.8	515	13	US-08-936-487-52
4	100	49.8	515	26	US-60-042-611-52
5	100	49.8	515	26	US-60-042-985-52
6	100	49.8	1144	26	US-60-439-453-167
7	100	49.8	2144	26	US-60-141-323-3

8	4.0	220	22	US-09-823-823-6	Sequence 6, Appli
9	4.0	220	22	US-09-823-823-6	Sequence 6, Appli
10	4.0	411	18	US-09-489-030A-7912	Sequence 7912, Ap
11	4.0	445	22	US-09-897-516-5029	Sequence 5029, Ap
12	4.0	445	22	US-60-215-161-5029	Sequence 5029, Ap
13	4.0	457	19	US-09-543-681A-6481	Sequence 6481, Ap
14	4.0	460	22	US-09-847-516-7795	Sequence 7795, Ap
15	4.0	460	22	US-60-215-161-7795	Sequence 7795, Ap
16	4.0	470	16	US-09-252-691-7148	Sequence 7148, Ap
17	4.0	470	16	US-09-252-691-7148	Sequence 7148, Ap
18	4.0	473	18	US-09-489-039A-9854	Sequence 9854, Ap
19	4.0	480	19	US-09-543-681A-4532	Sequence 4532, Ap
20	4.0	483	16	US-09-252-691-8532	Sequence 8532, Ap
21	4.0	483	16	US-09-252-691-8532	Sequence 8532, Ap
22	4.0	483	16	US-09-252-691-8532	Sequence 8532, Ap
23	4.0	483	16	US-09-252-691-8532	Sequence 8532, Ap
24	4.0	492	16	US-09-252-691-8532	Sequence 8532, Ap
25	7	36	1	PCT-US01-006663-4446	Sequence 4446, A
26	7	36	1	US-09-864-761-4446	Sequence 4446, A
27	7	36	1	US-08-417-872A-32	Sequence 32, Appl
28	7	36	1	US-09-879-967-79	Sequence 79, Appl
29	7	36	1	US-09-758-451-174	Sequence 174, Appl
30	7	36	1	US-16-020-386-2910	Sequence 2910, A
31	7	36	1	US-09-751-181-38	Sequence 38, Appl
32	7	36	1	US-60-170-373-2898	Sequence 2898, Ap
33	7	36	1	US-09-252-691-9390	Sequence 9390, Ap
34	7	36	1	US-09-252-691-9390	Sequence 9390, Ap
35	7	36	1	US-09-760-479-748	Sequence 748, App
36	7	36	1	US-09-867-550-624	Sequence 524, App
37	7	36	1	US-60-170-373-4141	Sequence 4141, Ap
38	7	36	1	US-09-489-039A-7932	Sequence 7932, Ap
39	7	36	1	US-09-641-377-184	Sequence 184, App
40	7	36	1	US-09-519-232-40	Sequence 40, Appl
41	7	36	1	US-09-339-972-57	Sequence 57, Appl
42	7	36	1	US-09-675-784A-13298	Sequence 13298, A
43	7	36	1	PCT-US02-03987-11461	Sequence 11461, A
44	7	36	1	US-09-815-242-11461	Sequence 11461, A
45	7	36	1	US-10-072-851-11461	Sequence 11461, A

ALIGNMENTS

RESULT 1
US-09-603-665-5
; Sequence 5, Application US/09603665
; GENERAL INFORMATION:
; APPLICANT: Barry, Caroline
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen-Akennine, Annick
; TITLE OF INVENTION: A NOVEL RAP28 GENE AND PROTEIN
; FILE REFERENCE: GENSET 063AUS
; CURRENT APPLICATION NUMBER: US/09/603,665
; PRIOR FILING DATE: 2000-06-23
; PRIOR FILING NUMBER: US 60/141,323
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/176,880
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent.pm
; SEQ ID NO 5
; LENGTH: 2144
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 1694
; OTHER INFORMATION: Xaa-Set or Asn
; NAME/KEY: VARIANT
; LOCATION: 1854
; OTHER INFORMATION: Xaa-Ala or Val
; NAME/KEY: VARIANT

EA Vazquez Boland L-A.; Voss H.; Wehland J.; Gossart P.;
EL "Comparative genomics of *Listeria* species";
RL Science 294:849-852(2001).
DR EMBL: AF56168; CAC56541.1; -
LK Listilist; LIN01510; -
KW Complete genome.
SQ SEQUENCE 180 AA; 29757 MW; 964E2A1A7332F22 CRC64;

Query Match 0.44; Score 8; DB 16; Length 180;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1269 DGGKTPKD 1276
|||||||
|b 129 DGGKTPKD 136

Search completed: July 2, 2002, 15:42:19
Job time: 457 sec

```

RA Kawayabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yushirawa T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.,
RT "Complete genome sequence of an aerobic thermocacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7.",
RL DNASIS: 8:123-140(2001).
KW EMBL: AF000982; HA65282.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 167 AA, 19252 MW, 2C1CA6DAE1502FA3 CRC64;

Query Match 0.4%; Score 8; DB 17; Length 167;
Best local Similarity 100.0%; Prod. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 EHRRLPIL 1399
Db 80 EHRRLPIL 87

RESULT 13
Q96KC2 PRELIMINARY; PRT; 179 AA.
AC Q96KC2;
DT 01-DEC-2001 (TRMBLrel. 19, Created)
DT 01-DEC-2001 (TRMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE CDNA FLJ114369 F1S, CLONE HEMPA1001174, HIGHLY SIMILAR TO
DE ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 5.
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Primates, Catartini, Homiidae, Homo.
OX NCBI_TaxID=9606;
RN [1]
RC TISSUE=EMHRYO, AND MAINLY HEAD;
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara P., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagaohari K., Masuho Y., Oshima A.;
RA "NHDO human cDNA sequencing project.";
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AK027275; HA855011.1; -.
SQ SEQUENCE 179 AA, 20374 MW, ECD74C4EALA90E33 CRC64;

Query Match 0.4%; Score 8; DB 4; Length 179;
Best local Similarity 100.0%; Prod. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1371 NVEEIVVK 1378
Db 51 NVEEIVVK 58

RESULT 14
Q9D4P0 PRELIMINARY; PRT; 179 AA.
AC Q9D4P0;
DT 01-JUN-2001 (TRMBLrel. 17, Created)
DT 01-JUN-2001 (TRMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE 4930587A11P1K PROTEIN.
GN 4930587A11P1K
OS Mus musculus (Mouse).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

PP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE=TESTIS;
RX MIML:21085660; PubMed:11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Balalov S., Casavani T.,
RA Pleischmann W., Gaasterland T., Gissi C., King R., Kochiwa H.,
RA Schmiel L.M., Straubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunda N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarolli J., Mombaerts P.,
RA Nordono P., King P., Hingwalji M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK016359; BAB30204.1; -.
DR HSSP: P32889; IHUR.
DP M3D: MGI:1921119; 4930587A11P1K.
DP InterPro: IPR001806, PASTrnsmfrmg.
DP InterPro: IPR002044; SArL_CTPBP
DP Pfam: PF000025; arf.1
DP PRINTS: PP00449; RASTRNSFRMG.
DP PRINTS: PP00128; RAS15TPBP.
DR SMART: SM00177; ARF; 1.
KW GTP-binding
SQ SEQUENCE 179 AA, 20374 MW, FCD74C4EALA90E33 CRC64;

Query Match 0.4%; Score 8; DB 11; Length 179;
Best local Similarity 100.0%; Prod. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1371 NVEEIVVK 1378
Db 51 NVEEIVVK 58

RESULT 15
Q92C81 PRELIMINARY; PRT; 180 AA.
AC Q92C81;
DT 01-DEC-2001 (TRMBLrel. 19, Created)
DT 01-DEC-2001 (TRMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE LIN1310 PROTEIN.
GN LIN1310
OS Listeria innocua.
OC Bacteria, Firmicutes, Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacteria.
OX NCBI_TaxID=1642;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEPOVAR 6A;
RX PubMed:11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Raquelero F., Berche P., Blocker J., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Ftsini H., Garcia-del Portillo E., Garrido P., Jackson D.,
RA Gantier L., Gobel W., Gomez-Lopez N., Hain T., Hain J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreff J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordisiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell P.,
RA Pemmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,

```

DR Problem: P00074; CM CSF: 1.
 DR SMART: SM00040; CSP2: 1.
 DR PROSITE: PS00702; CM_CSF: 1.
 FT VARIANT 60 60 V - I.
 SV SEQUENCE 144 AA: 16177 MW: 75515815A2F08 CF CRC64;

Query Match 0.4% Score 8; DB 6; Length 144;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 LLLIGTIVA 1412

|||||

DB 6 LLLIGTIVA 13

RESULT 9

ID Q9VZ05 PRELIMINARY; PRT: 146 AA.

AC Q9VZ05

DT 01-MAY-2000 (TrEMBLrel. 14, Created)

DT 01-MAY-2000 (TrEMBLrel. 14, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE HYPOTHETICAL 16.1 KDA PROTEIN.

GN PAR0008.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI_TaxID=29242;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN ORSAY;

RA Holig R;

PT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."

RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases

RL EMBL: AJ248283; CA449063.1; ..

RL Hypothetical protein: Complete proteome.

SV SEQUENCE 146 AA: 16092 MW: 718294147125801F CRC64;

Query Match

0.4% Score 8; DB 17; Length 146;

Best Local Similarity 100.0%; Pred. No. 59;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1041 LLEKIQE 1048

|||||

DB 101 LLEKIQE 108

RESULT 10

Q9LF40

ID Q9LF40 PRELIMINARY; PRT: 159 AA.

AC Q9LF40

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE HYPOTHETICAL 17.7 KDA PROTEIN.

GN T20K14.30.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=4702;

RN [1]

RP SEQUENCE FROM N.A.

RA Satoh S., Nakamura Y., Kaneko T., Kato T., Asanuma E., Kotani H.,

RA Tabata S., Mores H.W., Budd S., Lemcke K., Mayer K.F.X.,

RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases

RL EMBL: AJ241143; CA301741.1; ..

KW Hypothetical protein.
 SV SEQUENCE 159 AA: 17732 MW: FFA5D6503F3CFCF7 CRC64;

Query Match 0.4% Score 8; DB 10; Length 159;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 KPESLGKK 319

|||||

DB 9 KPESLGKK 16

RESULT 11

Q9JM20

ID Q9JM20 PRELIMINARY; PRT: 165 AA.

AC Q9JM20

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE ACTA (FRAGMENT).

GN ACTA.

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Listeria.

OX NCBI_TaxID=1639;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-L83;

XX MEDLINE-97342743; PubMed-979440;

RA Wiedmann M., Bruce J.L., Keating C., Johnson A.E., McDonough P.L.,

RA Batt C.A.;

RT "Ribotypes and virulence gene polymorphisms suggest three distinct Listeria monocytogenes lineages with differences in pathogenic potential."

RL Infect. Immun. 65:2707-2716(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-L83;

RA Wiedmann M., Bruce J.L., Keating C., Johnson A.E., McDonough P.L.,

RA Batt C.A.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.

RL EMBL: AF281893; AAF82597.1; ..

FT NON-TER 1

FT NON-TER 165

SV SEQUENCE 165 AA: 18272 MW: C0EFE780CEA860F1 CRC64;

Query Match

0.4% Score 8; DB 2; Length 165;

Best Local Similarity 100.0%; Pred. No. 66;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 FLPLRL 432

|||||

DB 152 FLPLRL 159

RESULT 12

Q975W6

ID Q975W6 PRELIMINARY; PRT: 167 AA.

AC Q975W6

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE HYPOTHETICAL PROTEIN ST0311.

GN ST0311.

OS Sulfolobus tokodaii.

OC Archaea; Ciliobacteriota; Sulfolobales; Sulfolobaceae; Sulfolobus.

OX NCBI_TaxID=111955;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-JCM 10545 / 7;

XX PubMed-11572479;

```

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 12.7 KDA PROTEIN.
GN DR0207.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales, Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eison J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
PA Pedson K.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Interhack T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.B., Salzberg S., Smith H.C., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL: AE001883; AAF09801.1;
DR FICP: PR0207;
KW Hypothetical protein, Complete proteome.
SQ SEQUENCE 123 AA, 12733 MW, 38936870.90840810 CRC64;

Query Match 0.4%; Score 8; DB 16; Length 123;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1757 LLSALAAL 1764
Dh 5 LLSALAAL 12
|||||||

RESULT 6
O57720 PRELIMINARY; PRT: 123 AA.
AC O57720;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DI 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 14.2 KDA PROTEIN PH1983.
GN PH1983.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE=98344147; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Hosoyama A., Nagai Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka K., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki T., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Poppe F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RI DNA Res. 5:55-76(1998).
DR EMBL: AP000007; BA331110.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 123 AA, 14212 MW, 177106136.00018 CRC64;

Query Match 0.4%; Score 8; DB 17; Length 123;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1655 KKKEGEEF 1662
Dh 1655 KKKEGEEF 1662
|||||||

```

```

Db 77 KKKEGEEF 84

RESULT 7
Q9YD74 PRELIMINARY; PRT: 123 AA.
AC Q9YD74;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DI 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 123AA LONG HYPOTHETICAL CYTIDINE DEAMINASE.
GN APE1038.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1;
RX MEDLINE=9910349; PubMed=10482466;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-oc K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kureta K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
RT "Complete genome sequence of an aerobic hyper thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RI DNA Res. 6:83-101(1999).
DR EMBL: A000060; MAA80023.1;
DR HSSP: F13652; ICTT.
DR InterPro: IPR002125; dCMP_cyt_deam.
DR Pfam: PF00383; dCMP_cyt_deam.1.
DR PROSITE: PS00604; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 123 AA; 13142 MW, 8FC1D7D8307F403C CRC64;

Query Match 0.4%; Score 8; DB 17; Length 123;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1031 LSQLPMA 1038
Dh 108 LSQLPMA 115
|||||||

RESULT 8
Q9GL44 PRELIMINARY; PRT: 144 AA.
AC Q9GL44;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DI 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR.
GN GM-CSF.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RA Hutchinson K.L., Villinger F., Miranda M.E., Ksiazek T.G.,
PA Peters C.T., Pollin P.E.;
RT "Multiplex analysis of cytokines in the sera of cynomolgus macaques
RT naturally infected with Ebola (Reston).";
PL Submitted (AUG-2000) to the FMBI/GenBank/EMBL databases
DR EMBL: AY007376; AAG16626.1;
DR HSSP: P04141; 2GMF.
DR InterPro: IPR000773; GM-CSF.
DR Pfam: PF01109; GM-CSF; 1
DR PRINTS: PR00693; GMCSFACTOR.

```

Q9W467
ID Q9W467
AC Q9W467
DT 01 NOV 1999 (TRENDS) 12, Created
DT 01 NOV 1999 (TRENDS) 12, Last sequence update
DT 01 DEC 2001 (TRENDS) 19, Last annotation update
DE B42 MATING PHEROMONE PRECURSOR PHB2.2.
GN PHB2.2.
OS Coprinus cinereus (Turkey cap fungus).
OA Fungi: Basidiomycota: Hymenogymetes: Homobasidiomycetes.
OA Agaricales: Coprinaceae: Coprinus.
OX NCBI TaxID=5146;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9926241; PubMed=10329522;
KA Giesbrecht N.S., Brown A.J., Dowell S.J., Casselton L.A.;
KI "A constitutively active G-protein-coupled receptor causes mating
self-compatibility in the mushroom *Coprinus citrinus*."
BL EMBO J. 18:2756-2761(1999).
EN [2]
FP SEQUENCE FROM N.A.
FX MEDLINE=20221720; PubMed=10707075;
KA Hallsall J.R., Milner M.J., Casselton L.A.;
KI "Three Subfamilies of Pheromone and Receptor Genes Generate Multiple B
Mating Specificities in the Mushroom *Coprinus citrinus*."
BL Genet Res 154:1115-1127(2000).
DR EMBL: AF146522; AAD27448.1;
SQ SEQUENCE 59 AA: 6462 MW: 96A459A2R4540695 CR664;

Query Match 0.4% Score 8; DB 3; Length 59;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 976 DLATLEE 98;
DB 7 DLATLEE 14

RESULT 4
Q9W467
ID Q9W467
AC Q9W467
DT 01 MAY 2000 (TRENDS) 13, Created
DT 01 MAY 2000 (TRENDS) 13, Last sequence update
DT 01 MAY 2000 (TRENDS) 13, Last annotation update
DE G11462 PROTEIN.
GN G11462.
OS Drosophila melanogaster (Fruit fly).
OA Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;
OA Protista: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha;
OA Ephydrioidae: Drosophilidae: Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
FX MEDLINE=2019609; PubMed=10741132;
KA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Geayne J.D.,
KA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallo R.F.,
KA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
KA Sutton G., Workman J.R., Yandell M.D., Zhang Q., Chen L.X.,
KA Branton R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
KA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
KA Abril J.F., Adabany A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
KA Balow R.M., Basu A., Haxendale J., Hayraktarolu I., Heasley E.M.,
KA Benson K.Y., Benos P.V., Bernin B.P., Bhandari D., Bolshakov S.,
KA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
KA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
KA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
KA de Pablos H., Belcher A., Bent Z., Mays A.D., Jew I., Dietz S.M.,
KA Dodson K., Doup L.E., Downes M., Druhan-Kocha S., Dunkov B.C., Dunn P.,
KA Durbin K.J., Evangelista C.G., Ferraz C., Ferreira S., Fleischman W.,
KA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iredam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel H.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mushrefi A.,
RA Mount S.M., Moy M., Murphy H., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Rossier D.K., Tachibana J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reineri K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skopski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.G.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."
FX Science 287:2185-2195(2000).
DR EMBL: AF003435; AAP46091.1;
RP FlyBase: FBq0029815; G11462.
SQ SEQUENCE 107 AA: 12104 MW: AE21FB9EF5771b5a CR664;

Query Match 0.4% Score 8; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 707 ILSVLVSC 714
DB 75 ILSVLVSC 82

RESULT 4
Q9UXU1
ID Q9UXU1
AC Q9UXU1
DT 01 MAY 2000 (TRENDS) 13, Created
DT 01 MAY 2000 (TRENDS) 13, Last sequence update
DT 01 JUN 2000 (TRENDS) 14, Last annotation update
DE HYPOTHETICAL 12.5 KDA PROTEIN.
GN PAB179.
OS Pyrococcus abyssi.
OA Archaea: Euryarchaeota: Thermococcales: Thermococcaceae: Pyrococcus.
OX NCBI TaxID 29292;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ORSAY;
KA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution."
RL submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: A7248288; CAB50672.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA: 12527 MW: 170274A4A9A59C1 CR664;

Query Match 0.4% Score 8; DB 17; Length 108;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1655 KKKKEEEE 1662
DB 62 KKKKEEEE 69

RESULT 5
Q9RXU9
ID Q9RXU9
AC Q9RXU9;
PRELIMINARY: PRI: 125 AA.

GenCore version 4.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein protein search, using sw model

Run On: July 2, 2002, 15:46:22, Search time: 68.45 seconds
(without alignments)
5418.576 Million cell updates/sec

Title: US 09 603-665-5
Perfect score: 2144
Sequence: 1 MTS:AAQGLQRIALFQSDASH.....CCKTKTQQUTVLGRTNQSYF 2144

Scoring table:

Gapop 60.0, Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size: 6

Total number of hits satisfying chosen parameters: 18454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

SPRMBH_19:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	5.9	349	4 Q96ES5	Q96ES5 homo sapien
2	8	0.4	59	3 Q9Y8C3	Q9Y8C3 coprinus ci
3	8	0.4	107	5 Q9W467	Q9W467 drosophila
4	8	0.4	108	17 Q9XU01	Q9XU01 pyrococcus
5	8	0.4	123	16 Q9XU09	Q9XU09 delinococcus
6	8	0.4	123	17 Q57720	Q57720 pyrococcus
7	8	0.4	123	17 Q9YD74	Q9YD74 aeropyrum p
8	8	0.4	144	6 Q9GL44	Q9GL44 macaca mla
9	8	0.4	146	17 Q9V2D5	Q9V2D5 pyrococcus
10	8	0.4	159	10 Q9IF40	Q9IF40 arabidopsis
11	8	0.4	165	2 Q9JWZ0	Q9JWZ0 historia mo
12	8	0.4	167	17 Q975W6	Q975W6 sulfolobus
13	8	0.4	179	4 Q96K02	Q96K02 homo sapien
14	8	0.4	179	11 Q9D4P0	Q9D4P0 mus musculus
15	8	0.4	180	16 Q92C81	Q92C81 listeria in
16	8	0.4	188	16 Q9K6G3	Q9K6G3 bacillus ha

ALIGNMENTS

RESULT 1

Q96ES5 ID Q96ES5 PRELIMINARY; PRT; 349 AA.
AC Q96ES5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO HYPOTHETICAL PROTEIN FLJ10359.
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY, AND ADENOCARCINOMA;
RA Strassberg R.;
RL Submitted (JUL 2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011983; A011983.1; -.
SQ SEQUENCE 349 AA: 39921 MW: 44545.57777/497.04644;

Query Match: 5.9%, Score 127, DB 4, Length 349;
Best Local Similarity 100.0%; Pred No. 8 9e-115;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2018 ALMPPLVLDENRAGGGRKKPQRYTKHLTPCAGPSVAMADSLWKPINYLAKTRISS 2077
ID 223 ALMPPLVLDENRAGGGRKKPQRYTKHLTPCAGPSVAMADSLWKPINYLAKTRISS 282
QY 2078 PKVPFAALITVLAAEKIKENYIVLLPSTPLAFIMEDECFEVEIDCQKTKTQOLETVLG 2137
ID 283 PKVPFAALITVLAAEKIKENYIVLLPSTPLAFIMEDECFEVEIDCQKTKTQOLETVLG 342
QY 2138 KPLQSYF 2144
ID 343 EPLQSYF 349

RESULT 2

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

Wed Jul 3 09:02:12 2002

DB 5 KKKGGEE 12

RESULT 14

US-60-360-039-12910
 ; Sequence 12910, Application US/60360039
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Chen, Xianfeng
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)A
 ; CURRENT APPLICATION NUMBER: US/60/360,039
 ; CURRENT FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 12910
 ; LENGTH: 624
 ; TYPE: PRT
 ; ORGANISM: Aspergillus nidulans
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)-(624)
 ; OTHER INFORMATION: unsure at all Xaa locations
 US-60-360-039-12910

Query Match 0.4%; Score 8; DB 7; Length 624;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1400 VOLVDTLG 1407
 |||||
 Db 591 VOLVDTLG 598

RESULT 15

US-60-382-898-332
 ; Sequence 332, Application US/60382898
 ; GENERAL INFORMATION:
 ; APPLICANT: Hudson, Keith
 ; APPLICANT: et al.
 ; TITLE OF INVENTION: Plant Receptors and Ligands
 ; FILE REFERENCE: 1066P
 ; CURRENT APPLICATION NUMBER: US/60/382,898
 ; CURRENT FILING DATE: 2002-05-22
 ; NUMBER OF SEQ ID NOS: 1344
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 332
 ; LENGTH: 664
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-60-382-898-332

Query Match 0.4%; Score 8; DB 7; Length 664;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1655 KKKGGEE 1662
 |||||
 Db 567 KKKGGEE 574

Search completed: July 2, 2002, 15:41:07
 Job Time: 350 sec

us-09-603-665-5.olig6.rapn

Wed Jul 3 09:02:12 2002

US-10-105-299-4155

Query Match 0.4%; Score 8; DB 6; Length 179;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1371 NVEEIVVK 1378
| | | | | | | |
Db 51 NVEEIVVK 58

RESULT 7

US-10-155-881-32064
; Sequence 32064, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyva, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 32064
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Glycine max
US-10-155-881-32064

Query Match 0.4%; Score 8; DB 6; Length 205;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1304 ALLLLGTV 1311
| | | | | | | |
Db 18 ALLLLGTV 25

RESULT 8

US-10-155-881-18999
; Sequence 18999, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyva, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 18999
; LENGTH: 266
; TYPE: PRT
; ORGANISM: glycine max
US-10-155-881-18999

Query Match 0.4%; Score 8; DB 6; Length 266;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1304 ALLLLGTV 1311
| | | | | | | |
Db 79 ALLLLGTV 86

; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 383
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 299
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: TAX43 human tax interaction protein 43 PDZ domain
; OTHER INFORMATION: 1
US-09-688-017-299

Query Match 0.4%; Score 8; DB 5; Length 78;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 QTQALYVG 897
| | | | | | | |
Db 42 QTQALYVG 49

RESULT 5

US-10-116-355-211
; Sequence 211, Application US/10116355
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins, and Antibodies
; FILE REFERENCE: PALLIC
; CURRENT APPLICATION NUMBER: US/10/116,355
; CURRENT FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 763
; Prior Application removed - See file Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 211
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-355-211

Query Match 0.4%; Score 8; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 LLLLLGTV 1312
| | | | | | | |
Db 80 LLLLLGTV 87

RESULT 6

US-10-105-299-4155
; Sequence 4155, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15187
; Prior Application removed - See file Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4155
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo sapiens

1 PRIOR APPLICATION NUMBER: US 60/160,860
2 PRIOR FILING DATE: 1999-10-21
3 PRIOR APPLICATION NUMBER: US 60/162,498
4 PRIOR FILING DATE: 1999-10-29
5 PRIOR APPLICATION NUMBER: US 60/170,454
6 PRIOR FILING DATE: 1999-12-13
7 PRIOR APPLICATION NUMBER: US 60/176,195
8 PRIOR FILING DATE: 2000-01-14
9 PRIOR APPLICATION NUMBER: US 60/182,296
10 PRIOR FILING DATE: 2000-02-14
11 PRIOR APPLICATION NUMBER: US 60/196,267
12 PRIOR FILING DATE: 2000-04-11
13 PRIOR APPLICATION NUMBER: US 60/196,460
14 PRIOR FILING DATE: 2000-04-11
15 PRIOR APPLICATION NUMBER: US 60/196,527
16 PRIOR FILING DATE: 2000-04-11
17 PRIOR APPLICATION NUMBER: US 60/196,528
18 PRIOR FILING DATE: 2000-04-11
19 PRIOR APPLICATION NUMBER: US 09/547,276
20 PRIOR FILING DATE: 2000-05-12
21 PRIOR APPLICATION NUMBER: US 09/569,525
22 PRIOR FILING DATE: 2000-05-12
23 PRIOR APPLICATION NUMBER: US 09/570,118
24 PRIOR FILING DATE: 2000-05-12
25 PRIOR APPLICATION NUMBER: US 09/570,364
26 PRIOR FILING DATE: 2000-05-12
27 PRIOR APPLICATION NUMBER: US 09/588,017
28 PRIOR FILING DATE: 2000-10-13
29 NUMBER OF SEQ ID NOS: 544
30 SOFTWARE: FastSeq for Windows Version 3.0
31 SEQ ID NO: 276
32 LENGTH: 72
33 TYPE: PRT
34 ORGANISM: Artificial Sequence
35 FEATURE:
36 OTHER INFORMATION: TAX ID 43 p02 domain 1
US-09-710-059-276

Query Match 0.4% Score 8; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 QTOALYVG 897
|||||
DB 42 QTOALYVG 49

RESULT 3
US-09-722-069-276
1 Sequence 276, Application US/09/722,069
2 GENERAL INFORMATION:
3 APPLICANT: Lu, Peter S.
4 APPLICANT: Rabinowitz, Joshua D.
5 APPLICANT: Schweizer, Johannes
6 APPLICANT: Arbor Vita Corporation
7 TITLE OF INVENTION: Molecular Interactions in Hematopoietic
8 TITLE OF INVENTION: Cells
9 FILE REFERENCE: 020054-001122US
10 CURRENT APPLICATION NUMBER: US/09/722,069
11 CURRENT FILING DATE: 2000-11-24
12 PRIOR APPLICATION NUMBER: US 60/134,114
13 PRIOR FILING DATE: 1999-05-14
14 PRIOR APPLICATION NUMBER: US 60/134,117
15 PRIOR FILING DATE: 1999-05-14
16 PRIOR APPLICATION NUMBER: US 60/134,118
17 PRIOR FILING DATE: 1999-05-14
18 PRIOR APPLICATION NUMBER: US 60/160,860
19 PRIOR FILING DATE: 1999-10-21
20 PRIOR APPLICATION NUMBER: US 60/162,498
21 PRIOR FILING DATE: 1999-10-29
22 PRIOR APPLICATION NUMBER: US 60/170,453
23 PRIOR FILING DATE: 1999-12-13

1 PRIOR APPLICATION NUMBER: US 60/176,195
2 PRIOR FILING DATE: 2000-01-14
3 PRIOR APPLICATION NUMBER: US 60/182,296
4 PRIOR FILING DATE: 2000-02-14
5 PRIOR APPLICATION NUMBER: US 60/196,267
6 PRIOR FILING DATE: 2000-04-11
7 PRIOR APPLICATION NUMBER: US 60/196,460
8 PRIOR FILING DATE: 2000-04-11
9 PRIOR APPLICATION NUMBER: US 60/196,527
10 PRIOR FILING DATE: 2000-04-11
11 PRIOR APPLICATION NUMBER: US 60/196,528
12 PRIOR FILING DATE: 2000-04-11
13 PRIOR APPLICATION NUMBER: US 09/547,276
14 PRIOR FILING DATE: 2000-04-11
15 PRIOR APPLICATION NUMBER: US 09/569,525
16 PRIOR FILING DATE: 2000-05-12
17 PRIOR APPLICATION NUMBER: US 09/570,118
18 PRIOR FILING DATE: 2000-05-12
19 PRIOR APPLICATION NUMBER: US 09/570,364
20 PRIOR FILING DATE: 2000-05-12
21 PRIOR APPLICATION NUMBER: US 09/588,017
22 PRIOR FILING DATE: 2000-10-13
23 PRIOR APPLICATION NUMBER: US 09/710,059
24 PRIOR FILING DATE: 2000-11-10
25 NUMBER OF SEQ ID NOS: 544
26 SOFTWARE: FastSeq for Windows Version 3.0
27 SEQ ID NO: 276
28 LENGTH: 72
29 TYPE: PRT
30 ORGANISM: Artificial Sequence
31 FEATURE:
32 OTHER INFORMATION: TAX ID 43 p02 domain 1
US-09-722-059-276

Query Match 0.4% Score 8; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 QTOALYVG 897
|||||
DB 42 QTOALYVG 49

RESULT 4
US-09-688-017-299
1 Sequence 299, Application US/09/688,017
2 GENERAL INFORMATION:
3 APPLICANT: Lu, Peter S.
4 APPLICANT: Rabinowitz, Joshua D.
5 APPLICANT: Schweizer, Johannes
6 APPLICANT: Arbor Vita Corporation
7 TITLE OF INVENTION: Molecular Interactions in Hematopoietic
8 TITLE OF INVENTION: Cells
9 FILE REFERENCE: 020054-001100US
10 CURRENT APPLICATION NUMBER: US/09/688,017
11 CURRENT FILING DATE: 2000-10-13
12 PRIOR APPLICATION NUMBER: US 60/134,114
13 PRIOR FILING DATE: 1999-05-14
14 PRIOR APPLICATION NUMBER: US 60/134,117
15 PRIOR FILING DATE: 1999-05-14
16 PRIOR APPLICATION NUMBER: US 60/134,118
17 PRIOR FILING DATE: 1999-05-14
18 PRIOR APPLICATION NUMBER: US 60/160,860
19 PRIOR FILING DATE: 1999-10-21
20 PRIOR APPLICATION NUMBER: US 60/162,498
21 PRIOR FILING DATE: 1999-10-29
22 PRIOR APPLICATION NUMBER: US 60/170,453
23 PRIOR FILING DATE: 1999-12-13
24 PRIOR APPLICATION NUMBER: US 60/176,195
25 PRIOR FILING DATE: 2000-01-14
26 PRIOR APPLICATION NUMBER: US 60/182,296
27 PRIOR FILING DATE: 2000-02-14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 15:35:17 : Search time 43 seconds
(without alignments)
5036.685 Million cell updates/sec

Title: US-09-603-665-5
Perfect score: 2144
Sequence: 1 MSLAQQQLRIALVQSDASLCORTIOQLFW:GHFLOSYP 2144

Scoring table: 0.150
Gapop 60 0 , Gapext 60 0

Searched: 307805 seqs, 101015595 residues

Word size : 6

Total number of hits satisfying chosen parameters: 10746

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents-AA_Ncw.*
1: /cgn2_5/prodata/1/paa/pct_NEW_COMB.pep.*
2: /cgn2_5/prodata/1/paa/US05_NEW_COMB.pep.*
3: /cgn2_5/prodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_5/prodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_5/prodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_5/prodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_5/prodata/1/paa/US16_NEW_COMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	10	0.5	505	b	US-10-141-675-2
2	8	0.4	72	5	US-09-710-059-276
3	8	0.4	72	5	US-09-710-059-276
4	8	0.4	78	5	US-09-688-017-294
5	8	0.4	93	6	US-10-116-455-211
6	8	0.4	176	6	US-10-105-299-4155
7	8	0.4	205	6	US-10-155-881-32064
8	8	0.4	266	6	US-10-155-881-18999
9	8	0.4	467	5	US-09-540-209-8924
10	8	0.4	543	7	US-09-360-039-20635
11	8	0.4	554	5	US-09-573-655B-1077
12	8	0.4	603	5	US-09-935-625-9919
13	8	0.4	606	5	US-09-935-625-9918
14	8	0.4	624	7	US-60-360-039-12910
15	8	0.4	664	7	US-60-382-898-332
16	8	0.4	918	7	US-60-360-039-18355
17	8	0.4	1083	5	US-09-935-625-25456
18	8	0.4	1192	5	US-09-935-625-25454
19	8	0.4	1215	5	US-09-935-625-25454
20	8	0.4	1357	5	US-09-573-655B-2227
21	8	0.4	2109	6	US-60-360-039-18355
22	7	0.3	15	6	US-10-134-519-1
23	7	0.3	57	6	US-10-155-881-6943
24	7	0.3	58	5	US-09-620-393B-1821
25	7	0.3	62	5	US-09-620-393B-2792
26	7	0.3	70	5	US-09-620-393B-3023

ALIGNMENTS

RESULT 1
US-10-141-675-2
: Sequence 2, Application US/10141675
: GENERAL INFORMATION:
: APPLICANT: Curtis, Rory A. J.
: TITLE OF INVENTION: 63751, Human Signal Transporter Family
: FILE REFERENCE: MP101-095PIRM
: CURRENT APPLICATION NUMBER: US/10/141-675
: PRIOR FILING DATE: 2002-05-08
: PRIOR APPLICATION NUMBER: 60/289,895
: PRIOR FILING DATE: 2001-01-09
: NUMBER OF SEQ. IN NOS: 7
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 565
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-141-675-2

Query Match 0.5%, Score 10; DB 6; Length 565;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1472 KLPEKEETI 1481
Db 541 KLPEKEETI 550

RESULT 2
US-09-710-059-276
: Sequence 276, Application US/09710059
: GENERAL INFORMATION:
: APPLICANT: Lu, Peter S.
: APPLICANT: Rabinowitz, Joshua D.
: APPLICANT: Schweitzer, Johannes
: APPLICANT: Arbor Vita Corporation
: TITLE OF INVENTION: Molecular Interactions in Hematopoietic
: FILE REFERENCE: 020054, 00112005
: CURRENT APPLICATION NUMBER: US/09/710,059
: CURRENT FILING DATE: 2000-10-11
: PRIOR APPLICATION NUMBER: US 60/134,114
: PRIOR FILING DATE: 1999-05-14
: PRIOR APPLICATION NUMBER: US 60/134,117
: PRIOR FILING DATE: 1999-05-14
: PRIOR APPLICATION NUMBER: US 60/134,118
: PRIOR FILING DATE: 1999-05-14

Sequence 3023, Ap
Sequence 8069, Ap
Sequence 30859, A
Sequence 30858, A
Sequence 941, App
Sequence 5771, Ap
Sequence 3021, Ap
Sequence 2562, Ap
Sequence 6607, Ap
Sequence 31, Appli
Sequence 4689, Ap
Sequence 4716, Ap
Sequence 5243, Ap
Sequence 4340, Ap
Sequence 6606, Ap
Sequence 685, App
Sequence 9, Appli
Sequence 10675, A
Sequence 30857, A

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

QY 684 VEDLISVG 691
DB 80 VEDLISVG 87

RESULT 15

S44768
C29E4.7 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 14-Sep-1994 #sequence_revision 12 May 1995 #text_change 23-Mar 2001
C:Accession: S44768
R:Wilson, R.
submitted to the EMBL Data Library, September 1993
A:Description: Sequence of the C. elegans cosmid C29E4.
A:Reference number: S44738
A:Accession: S44768
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-250 <WIL>
A:Cross-references: EMBL:123651; NID:g388575; PID:g388579
C:Genetics:
A:Introns: 49/2; 122/3; 165/1

Query Match 0.4%; Score 8; DB 2; Length 250;
Best local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 VEDLISVG 691
DB 80 VEDLISVG 87

Search completed: July 2, 2002, 15:37:28
Job time: 236 sec

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <S10>
A:Cross-references: GH:AF001519; GH:HA060004; NID:q10176109; PIDN:HA07485.1; GSPDB:GN00148
A:Experimental source: strain G-125
C:Genetics:
A:Gene: BH4766
C:Superfamily: Bacillus subtilis hypothetical protein ywlg

Query Match 0.48; Score 8; DB 2; Length 188;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2140 CEELIVLG 2147
|||||||
DB 6 QQLRIVLG 13

RESULT 11
A87259
hypothetical protein G0082 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20 Apr 2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87259
R:Neerman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; Laub, M.T.; DeBoy, R.J.; Dodson, R.J.; Parkin, A.S.; Swift, M.L.; Haft, D.H.; Kolonits, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87259
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <S10>
A:Cross-references: GH:AF000633; NID:q1421181; PIDN:AAK22069.1; GSPDB:GN00148
C:Genetics:
A:Gene: G0082

Query Match 0.48; Score 8; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 LLSRDEVA 27
|||||||
DB 196 LLSRDEVA 203

RESULT 12
G82276
hypothetical protein V0811 [imported] - Vibrio cholerae (strain N1691 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18 Aug 2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: G82276
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Churchland, J.E.; Ermolaeva, M.; Vamathevan, J.; Hass, S.; Qin, H.; Dragol, I.; Sellers, J.L.; Rhee, M.; McKelvey, J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82045; MUID:20406843
A:Accession: G82276
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <HEL>
A:Cross-references: GH:AE001466; GH:AE0014852; NID:q0655259; PIDN:AAF93375.1; GSPDB:GN00148
A:Experimental source: serogroup O1; strain N1691; biotype El Tor
C:Genetics:
A:Gene: V0811
A:Map position: 1

Query Match 0.48; Score 8; DB 2; Length 234;

Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1038 ABOLLEKI 1045
|||||||
DB 120 ABOLLEKI 127

RESULT 13
140627
probable transcription initiation factor sigma E - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 12 Aug 1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C:Accession: 140627; S34309
R:Wood, J.; Sass, C.; Bennett, G.N. Gene 153, 89-92, 1995
A:Title: Sequence and arrangement of genes encoding sigma factors in Clostridium acetobutylicum
A:Reference number: 140626; MUID:q5189110
A:Accession: 140627
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-235 <RES>
A:Cross-references: EMBL:U07420; NID:q705344; PIDN:AAK43309.1; PIDN:q460971
R:Saier, W.; Treuner, A.; Buchholz, M.; Duerre, P. submitted to the EMBL Data Library, June 1993
A:Description: Sigma factor homologous genes in C. acetobutylicum.
A:Reference number: S34306
A:Accession: S34309
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 46-146, 148-235 <SA0>
A:Cross-references: EMBL:Z24079
C:Genetics:
A:Gene: sigE
C:Superfamily: transcription initiation factor sigma K; transcription initiation factor E
C:Keywords: DNA binding; sigma factor; transcription initiation
E:660-235/Domain: transcription initiation factor sigma K; transcription initiation factor E

Query Match 0.48; Score 8; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 684 VEDLISVG 691
|||||||
DB 80 VEDLISVG 87

RESULT 14
B97109
DNA-dependent RNA polymerase sigma chain [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: B97109
R:Mulliq, J.; Brelton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R. J. Bacteriol. 183, 4833-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359425; PMID:12149325
A:Accession: B97109
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-235 <K09>
A:Cross-references: GH:AF001437; PIDN:AAK7461.1; PIDN:q15024659; GSPDB:GN00148
A:Experimental source: Clostridium acetobutylicum AUCB24
C:Genetics:
A:Gene: CAC1695
C:Superfamily: transcription initiation factor sigma K; transcription initiation factor E

Query Match 0.48; Score 8; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 LIAAGCTGA 1312
 |||||
 Db 6 LIAAGCTGA 13

RESULT 6
 H75201
 hypothetical protein PAB0088 - Pyrococcus abyssi (strain Orsay)
 C:Species: Pyrococcus abyssi
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 04-Sep-2000
 C:Accession: H75201
 P:Anonymous; Genoscope
 A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
 A:Reference number: A75001
 A:Accession: H75201
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <RAW>
 A:Cross references: GH:AJ248-83; GH:AJ296-85; NID:J5457433; PIDD:CAR49063.1, PIDD:J51495
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: PAB0088
 C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0088

Query Match 0.4%; Score 8; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1041 LAKKIQKE 1048
 |||||
 Db 101 LAKKIQKE 108

RESULT 7
 T51520
 hypothetical protein T20K14_30 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse ear cross)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
 C:Accession: T51520
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
 submitted to the Protein Sequence Database, August 2000
 A:Reference number: Z25394
 A:Accession: T51520
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <SAT>
 A:Cross references: EMBL:AI391143
 A:Experimental source: cultivar Columbia; PAC clone T20K14
 C:Genetics:
 A:Map position: 5
 A:Introns: 11/1
 A>Note: T20K14_30

Query Match 0.4%; Score 8; DB 2; Length 159;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 KPSIGKK 319
 |||||
 Db 9 KPSIGKK 16

RESULT 8
 A61596
 signal peptidase I homolog lin1310 [imported] Listeria innocua (strain Clip1262)
 C:Species: Listeria innocua
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: A61596

R:Glaser, P.; Frangeul, L.; Buchrieser, G.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkak, G.; Madueno, E.; Maitournam, A.;
 Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID-21537279; PMID:11679669
 A:Accession: A61596
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-180 <GLA>
 A:Cross references: GB:A592022, PIDD:CA9541.1, PIDD:J41478, GSPDK:GN00177
 A:Experimental source: strain Clip1262
 C:Genetics:
 A:Gene: lin1310
 C:Superfamily: signal peptidase I sips

Query Match 0.4%; Score 8; DB 2; Length 180;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1269 DGGKIPKD 1276
 |||||
 Db 129 DGGKIPKD 136

RESULT 9
 AG1233
 signal peptidase I homolog lmo1271 [imported] - Listeria monocytogenes (strain EGD-e)
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: AG1233
 R:Glaser, P.; Frangeul, L.; Buchrieser, G.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkak, G.; Madueno, E.; Maitournam, A.;
 Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID-21537279; PMID:11679669
 A:Accession: AG1233
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-180 <GLA>
 A:Cross references: GH:NC_003210; PIDD:CA93349.1; PIDD:J15410687; GSPDK:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo1271
 C:Superfamily: signal peptidase I sips

Query Match 0.4%; Score 8; DB 2; Length 180;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1269 DGGKIPKD 1276
 |||||
 Db 129 DGGKIPKD 136

RESULT 10
 F84120
 hypothetical protein BH3766 [imported] Bacillus halodurans (strain C 125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: F84120
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Macno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: AF0650; MUID 2051562; PMID:11958132
 A:Accession: F84120

Query Match 0.4% Score 8; DB 1; Length 123;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1655 KKKEGEE 1662
 16 1111111
 1b 77 KKKEGEE 84

RESULT 4
 672702
 Probable cytidine deaminase AFE038 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C>Date: 20 Aug 1999 #sequence_revision 20 Aug 1999 #text_change 20 Jun 2000
 C:Accession: G72702
 C:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kuboh, Y.; Yamazaki, J.; Kana, R.; 6, 84-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
 A:Reference number: A72450; M010:9410339
 A:Accession: G72702
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-123 <KAW>
 A:Cross-references: EMBL:AF000060; NID:q5104188; PID:BAA80023.1; PID:q5104708
 A:Experimental source: Strain K1
 C:Genetics:
 A:Gene: APE038
 C:Superfamily: cytidine deaminase

Query Match 0.4% Score 8; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1041 LSQLLPMA 1048
 11111111
 1b 108 LSQLLPMA 115

RESULT 4
 R75546
 Hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: deinococcus radiodurans
 C>Date: 04 Dec 1999 #sequence_revision 03 Dec 1999 #text_change 31 Mar 2000
 C:Accession: R75546
 C:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, T.V.; Lam, P.; McDonald, L.; Utterback, T.; Zolowski, C.; M.; Smith, B.G.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radiotolerant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; M010:20036836
 A:Accession: R75546
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-124 <WHI>
 A:Cross-references: GB:AF001883; NID:q6457878; PID:AAF0980.1; PID:q645788
 A:Experimental source: Strain R1
 C:Genetics:
 A:Gene: DR0207
 A:Map position: 1

Query Match 0.4% Score 8; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1757 LLSALAAAL 1764
 11111111
 1b 11111111 LLSALAAAL 12

RESULT 5
 F01818M
 granulocyte-macrophage colony-stimulating factor precursor [validated] - human
 N:Alternate names: colony-stimulating factor; 2; GM-CSF
 C:Species: Homo sapiens (man)
 C>Date: 04 Dec 1986 #sequence_revision 04 Dec 1986 #text_change 08 Dec 2000
 C:Accession: G24636; A25169; A01853; A44175; J01090
 R:Miyatake, S.; Otsuka, T.; Yokota, T.; Lee, F.; Arat, K.
 EMBO J. 4, 2561-2568, 1985
 A:Title: Structure of the chromosomal gene for granulocyte macrophage colony stimulat
 A:Reference number: A91015; M010:86030234
 A:Accession: G24636
 A:Molecule type: DNA
 A:Residues: 1-144 <MIY>
 A:Cross-references: EMBL X03321, NID:q18147; PID:AAA98768.1; PID:q18148
 R:Kaushansky, K.; G'Harra, P.J.; Berknot, N.; Seegal, G.M.; Hagen, F.S.; Adamsen, J.W.
 Proc Natl Acad Sci U.S.A. 87, 3101-3105, 1986
 A:Title: Genomic cloning, characterization, and multilocus growth promoting activit
 A:Reference number: A59065; M010:86205844
 A:Accession: A59065
 A:Status: translated from GH/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-144 <RES>
 A:Cross-references: GB:MI3207; NID:q181147; PID:AAA98768.1; PID:q181148
 R:Gandrell, M.A.; Anderson, D.; Cerretti, D.P.; Price, V.; McKeredan, K.; Tushinski, P.
 Proc Natl Acad Sci U.S.A. 87, 6250-6254, 1985
 A:Title: Cloning, sequence, and expression of a human granulocyte/macrophage colony-s
 A:Reference number: A25169; M010:85298329
 A:Accession: A25169
 A:Molecule type: mRNA
 A:Residues: 1-144 <CAN>
 A:Cross-references: GB:MI1734; NID:q181149; PID:AAA52122.1; PID:q181150
 R:Lee, F.; Yokota, T.; Otsuka, T.; G'Harra, N.; Itoh, J.; Arat, K.; Kende
 Proc Natl Acad Sci U.S.A. 87, 4360-4364, 1985
 A:Title: Isolation of cDNA for a human granulocyte-macrophage colony-stimulating fact
 A:Reference number: A01853; M010:85242684
 A:Accession: A01853
 A:Molecule type: mRNA
 A:Residues: 1-144 <LEK>
 A:Cross-references: GB:MI1220; NID:q181364; PID:AAA52578.1; PID:q181364
 R:Wong, G.G.; Wittek, J.S.; Temple, P.A.; Wilkens, R.M.; Leary, A.C.; Luxenberg, D.P.;
 A.; Clark, S.C.
 Science 228, 810-815, 1985
 A:Title: Human GM-CSF: molecular cloning of the complementary DNA and purification of
 A:Reference number: A44175; M010:85218749
 A:Accession: A44175
 A:Molecule type: mRNA
 A:Residues: 1-115; T, 118-144 <WON>
 A:Cross-references: GB:MI0663; NID:q181145; PID:AAA52121.1; PID:q181146
 A:Note: parts of this sequence, including the amino end of the mature protein, were c
 R:Wen, D.Y.; Huang, B.R.; Cal, L.W.; Si, J.Y.
 Acta Biochim Biophys. Sin. 25, 651-655, 1993
 A:Title: Amplification of human granulocyte macrophage colony stimulating factor cDNA
 A:Reference number: J01090
 A:Accession: J01090
 A:Molecule type: protein
 A:Residues: 18-21, 23-96, 117, 98-144 <WEN>
 C:Genetics:
 A:Gene: GDR:CSF2
 A:Cross-references: GDB:119812; OMIM:138960
 A:Map position: 5q23.2-5q31.1
 A:Introns: 53/3; 67/3; 109/3
 C:Function:
 A:Description: stimulates the differentiation and proliferation of hematopoietic prog
 C:Superfamily: granulocyte macrophage colony stimulating factor
 C:Keywords: cytokine; glycoprotein; growth factor; macrophage; monomer; T cell
 F:1-17/Domain: signal sequence #status predicted <SIG>
 F:18-144/Product: granulocyte-macrophage colony-stimulating factor #status experiment
 F:14,54/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.4% Score 8; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 18;

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

RESULT 25
US-60-230-445-1772
; Sequence 1772, Application US/60230445
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000765
; CURRENT APPLICATION NUMBER: US/60/230,445
; CURRENT FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 3051
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1772
; LENGTH: 524
; TYPE: PRT
; ORGANISM: HUMAN
US-60-230-445-1772

Query Match 0.5%; Score 10; DH 26; Length 524;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 KLPEEKEETI 1481
|||||
DB 500 KLPEEKEETI 509

RESULT 26
US-60-230-445-1930
; Sequence 1930, Application US/60230445
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000765
; CURRENT APPLICATION NUMBER: US/60/230,445
; CURRENT FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 3051
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1930
; LENGTH: 524
; TYPE: PRT
; ORGANISM: HUMAN
US-60-230-445-1930

Query Match 0.5%; Score 10; DH 26; Length 524;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 KLPEEKEETI 1481
|||||
DB 500 KLPEEKEETI 509

Search completed: July 2, 2002, 16:22:13
Job time: 450 sec

APPLICANT: Harland, Lee
APPLICANT: Burford, Neil
APPLICANT: Greene, Barrie D.
APPLICANT: Sanjanwala, Madhu Sudan
APPLICANT: Baughn, Mariah R.
APPLICANT: Yao, Monique G.
APPLICANT: Yang, Junming
APPLICANT: Patterson, Chandra
APPLICANT: Gandhi, Ameena R.
APPLICANT: Hafalia, April
APPLICANT: Tribouley, Catherine M.
APPLICANT: Walla, Narinder K.
APPLICANT: Au-Young, Janice
APPLICANT: Walsh, Roddy
APPLICANT: Ramkumar, Jaya
APPLICANT: Lu, Yan
APPLICANT: Lu, Dyoung Aina M.
APPLICANT: Azimzai, Yalda
APPLICANT: Lal, Preeti
; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
; FILE REFERENCE: PI-0149 P
; CURRENT APPLICATION NUMBER: US/60/216,547
; CURRENT FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Inceyle ID No: 7475338CD1
US-60-216-547-7

Query Match 0.5%; Score 10; DH 26; Length 512;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 KLPEEKEETI 1481
|||||
DB 488 KLPEEKEETI 497

RESULT 24
US-09-822-863-2
; Sequence 2, Application US/09822863
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001002
; CURRENT APPLICATION NUMBER: US/09/822,863
; CURRENT FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Human
US-09-822-863-2

Query Match 0.5%; Score 10; DH 22; Length 516;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 KLPEEKEETI 1481
|||||
DB 492 KLPEEKEETI 501

: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
 : TITLE OF INVENTION: AND USES THEREOF
 : FILE REFERENCE: C1000451
 : CURRENT APPLICATION NUMBER: US/60/196,712
 : NUMBER OF SEQ ID NOS: 3846
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 3126
 : LENGTH: 64
 : TYPE: PRT
 : ORGANISM: HUMAN
 US 60 196 712-4126

Query Match 0.5% Score 10; DB 26; Length 54;

Best Local Similarity 100.0%; Pred. No. 1.7; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0;

QY 1472 KLPEKEETI 1481

DB 48 KLPEKEETI 57

RESULT 20

US 60 248 796-20694

: Sequence 20694, Application US/60/248796

: GENERAL INFORMATION:

: APPLICANT: Keith Weinstock et al

: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

: FILE REFERENCE: 107196.132

: CURRENT APPLICATION NUMBER: US/60/248,796

: NUMBER OF SEQ ID NOS: 1999-02-12

: SEQ ID NO 20694

: LENGTH: 241

: TYPE: PRT

: ORGANISM: Candida albicans

US 60 248 796-20694

Query Match

Best Local Similarity 100.0%; Pred. No. 5.8; Length 241;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1614 SVRRKALDLL 1623

DB 89 SVRRKALDLL 98

RESULT 21

US 60 096-409-20694

: Sequence 20694, Application US/60/96409A

: GENERAL INFORMATION:

: APPLICANT: Keith Weinstock et al

: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

: FILE REFERENCE: G10-0161

: CURRENT APPLICATION NUMBER: US/60/096,409A

: NUMBER OF SEQ ID NOS: 1998-08-13

: SEQ ID NO 20694

: LENGTH: 241

: TYPE: PRT

: ORGANISM: Candida albicans

US 60 096-409-20694

Query Match

Best Local Similarity 100.0%; Pred. No. 5.8; Length 241;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1614 SVRRKALDLL 1623

DB 89 SVRRKALDLL 98

RESULT 22

US 60 322-511-640

: Sequence 640, Application US/60322511

: GENERAL INFORMATION:

: APPLICANT: Tang, Y. Tom

: APPLICANT: Yang, Yonghong

: APPLICANT: Wang, Zhiwei

: APPLICANT: Weng, Gezhi

: APPLICANT: Ma, Yundong

: TITLE OF INVENTION: Novel Nucleic Acids and

: FILE REFERENCE: 807

: CURRENT APPLICATION NUMBER: US/60/322,511

: CURRENT FILING DATE: 2001-09-13

: PRIOR APPLICATION NUMBER: US 09/488,725

: PRIOR FILING DATE: 2000-01-21

: PRIOR APPLICATION NUMBER: US 09/552,417

: PRIOR FILING DATE: 2000-04-25

: PRIOR APPLICATION NUMBER: US 09/491,404

: PRIOR FILING DATE: 2000-01-25

: PRIOR APPLICATION NUMBER: US 09/496,914

: PRIOR FILING DATE: 2000-02-03

: PRIOR APPLICATION NUMBER: US 09/560,875

: PRIOR FILING DATE: 2000-04-27

: PRIOR APPLICATION NUMBER: US 09/515,126

: PRIOR FILING DATE: 2000-02-28

: PRIOR APPLICATION NUMBER: US 09/577,409

: PRIOR FILING DATE: 2000-05-18

: PRIOR APPLICATION NUMBER: US 09/519,705

: PRIOR FILING DATE: 2000-03-07

: PRIOR APPLICATION NUMBER: US 09/574,454

: PRIOR FILING DATE: 2000-05-19

: PRIOR APPLICATION NUMBER: US 09/540,217

: PRIOR FILING DATE: 2000-03-31

: PRIOR APPLICATION NUMBER: US 09/649,167

: PRIOR FILING DATE: 2000-08-23

: PRIOR APPLICATION NUMBER: US 09/552,929

: PRIOR FILING DATE: 2000-04-13

: PRIOR APPLICATION NUMBER: US 09/770,160

: PRIOR FILING DATE: 2001-01-25

: PRIOR APPLICATION NUMBER: US 09/577,408

: PRIOR FILING DATE: 2000-05-13

: NUMBER OF SEQ ID NOS: 998

: SOFTWARE: pt_FL_genes Version 5.0

: SEQ ID NO 640

: LENGTH: 486

: TYPE: PRT

: ORGANISM: Homo sapiens

US 60 322-511-640

Query Match

Best Local Similarity 100.0%; Pred. No. 11; Length 486;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 KLPEKEETI 1481

DB 462 KLPEKEETI 471

RESULT 23

US 60 216-547-7

: Sequence 7, Application US/60216547

: GENERAL INFORMATION:

: APPLICANT: Raumann, Brigitte K.

: APPLICANT: Thornton, Michael

: APPLICANT: Ding, Li

: APPLICANT: Yue, Henry

: APPLICANT: Tang, Y. Tom

```
; SEQ ID NO 2540
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-757-028-2550

Query Match
Best Local Similarity 100.0%; Score 72; DB 21; Length 135;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1766 KVVETLPHFISPYLEGILSOVHLEKITSEMGSASQANIRLTSLKKTATLAPRVLLPA 1825
Db 44 KVVETLPHFISPYLEGILSOVHLEKITSEMGSASQANIRLTSLKKTATLAPRVLLPA 103

QY 1826 IKKTYKQIEKNW 1837
Db 104 IKKTYKQIEKNW 115

RESULT 16
US-09-757-028-2541
; Sequence 2541, Application US/09757028
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM001
; CURRENT APPLICATION NUMBER: US/09/757,028
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/174,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 2660
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2541
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (78)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (100)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-757-028-2541

Query Match
Best Local Similarity 100.0%; Score 59; DB 21; Length 105;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1768 VETLPHFISPYLEGILSOVHLEKITSEMGSASQANIRLTSLKKTATLAPRVLLPAI 1826
Db 19 VETLPHFISPYLEGILSOVHLEKITSEMGSASQANIRLTSLKKTATLAPRVLLPAI 77

RESULT 17
US-09-757-028-1640
; Sequence 1640, Application US/09757028
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM001
; CURRENT APPLICATION NUMBER: US/09/757,028
; CURRENT FILING DATE: 2001-01-09
```

```
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 2660
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1640
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (41)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (161)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (166)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (167)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: N. Xaa equals any of the naturally occurring L-amino acids
US-09-757-028-1640
```

Query Match 2.5%; Score 53; DB 21; Length 170;
Best Local Similarity 100.0%; Pred. No. 4.8e-42; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 1641 FLKLVPLDIIAIVQKKKGEEQAINROTALYTLKILCKNFGAENPDPFVPL 1693
Db 42 FLKLVPLDIIAIVQKKKGEEQAINROTALYTLKILCKNFGAENPDPFVPL 94

RESULT 18
US-60-196-710-5862
; Sequence 5862, Application US/60196710
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000450
; CURRENT APPLICATION NUMBER: US/60/196,710
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7166
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5862
; LENGTH: 64
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-710-5862

Query Match 0.5%; Score 10; DB 26; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.7; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 1472 KLPERKEETI 1481
Db 48 KLPERKEETI 57

RESULT 19
US-60-196-712-3126
; Sequence 3126, Application US/60196712
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

1 APPLICANT: Rank, David R.
2 APPLICANT: Hanzel, David K.
3 APPLICANT: Chen, Wensheng
4 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
5 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
6 FILE REFERENCE: Asmeca-X-1
7 CURRENT APPLICATION NUMBER: US/09/864,761
8 PRIOR FILING DATE: 2001-05-23
9 PRIOR APPLICATION NUMBER: US 60/180,412
10 PRIOR FILING DATE: 2000-02-04
11 PRIOR APPLICATION NUMBER: US 60/207,456
12 PRIOR FILING DATE: 2000-05-26
13 PRIOR APPLICATION NUMBER: US 09/632,366
14 PRIOR FILING DATE: 2000-08-03
15 PRIOR APPLICATION NUMBER: GB 24263.6
16 PRIOR FILING DATE: 2000-10-04
17 PRIOR APPLICATION NUMBER: US 60/236,359
18 PRIOR FILING DATE: 2000-09-27
19 PRIOR APPLICATION NUMBER: PCT/US01/00666
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00667
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: PCT/US01/00664
24 PRIOR FILING DATE: 2001-01-30
25 PRIOR APPLICATION NUMBER: PCT/US01/00664
26 PRIOR FILING DATE: 2001-01-30
27 PRIOR APPLICATION NUMBER: PCT/US01/00665
28 PRIOR FILING DATE: 2001-01-30
29 PRIOR APPLICATION NUMBER: PCT/US01/00668
30 PRIOR FILING DATE: 2001-01-30
31 PRIOR APPLICATION NUMBER: PCT/US01/00663
32 PRIOR FILING DATE: 2001-01-30
33 PRIOR APPLICATION NUMBER: PCT/US01/00662
34 PRIOR FILING DATE: 2001-01-30
35 PRIOR APPLICATION NUMBER: PCT/US01/00661
36 PRIOR FILING DATE: 2001-01-30
37 PRIOR APPLICATION NUMBER: PCT/US01/00670
38 PRIOR FILING DATE: 2001-01-30
39 PRIOR APPLICATION NUMBER: US 60/234,687
40 PRIOR FILING DATE: 2000-09-21
41 PRIOR APPLICATION NUMBER: US 09/608,408
42 PRIOR FILING DATE: 2000-06-30
43 PRIOR APPLICATION NUMBER: US 09/774,204
44 PRIOR FILING DATE: 2001-01-29
45 NUMBER OF SEQ ID NOS: 49117
46 SOFTWARE: Atanmax Sequence Listing Engine vers. 1.1
47 SEQ ID NO: 47046
48 LENGTH: 77
49 TYPE: PPT
50 ORGANISM: Homo sapiens
51 FEATURE:
52 OTHER INFORMATION: MAP TO AC010807.2
53 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 1.3
54 OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL - 1.8
55 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL - 1.7
56 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 1.8
57 OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL - 1.5
58 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 1.8
59 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL - 1.3
60 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL - 1.7
61 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL - 1.4
62 OTHER INFORMATION: EST_HUMAN HIT: AL040338.1, EVALUATE 8.00e-37
63 OTHER INFORMATION: SWISSPROT HIT: P42945, EVALUATE 2.00e-03
64 US 09-864-761-47046

Query Match 3.6% Score 77: DB 22: Length 77:
Best Local Similarity 100.0% Pred. No. 8.9e-66;
Matches 77: Conserved 0: Mismatches 0: Indels 0: Gaps 0;
US 09-864-761-47046
1360 SDSGDSIEVSRNVEIVVKIISVFVDALPHVPEHRRPLPIIVQIVDTLGAEKFLWILLILL 1419
|||||

DB 1 SDSGDSIEVSRNVEIVVKIISVFVDALPHVPEHRRPLPIIVQIVDTLGAEKFLWILLILL 60
QY 1420 PQYVTKTVLAAAYGK 1436
|||||
DB 61 PQYVTKTVLAAAYGK 77
RESULT 14
US-60-236-359-19127
Sequence 19127, Application US/60236359
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
FILE REFERENCE: MOHORE-4P
CURRENT APPLICATION NUMBER: US/60/236,359
CURRENT FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
NUMBER OF SEQ ID NOS: 21709
SOFTWARE: Molecular Dynamics Sequence Listing Engine
SEQ ID NO: 19127
LENGTH: 77
TYPE: PPT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010807.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 1.3
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL - 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL - 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 1.8
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL - 1.5
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 1.8
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL - 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL - 1.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL - 1.4
OTHER INFORMATION: EST_HUMAN HIT: AL040338.1, EVALUATE 8.00e-37
OTHER INFORMATION: SWISSPROT HIT: P42945, EVALUATE 2.00e-03
US-60-236-359-19127

Query Match 3.6% Score 77: DB 26: Length 77:
Best Local Similarity 100.0% Pred. No. 8.9e-66;
Matches 77: Conserved 0: Mismatches 0: Indels 0: Gaps 0;

QY 1360 SDSGDSIEVSRNVEIVVKIISVFVDALPHVPEHRRPLPIIVQIVDTLGAEKFLWILLILL 1419
|||||
DB 1 SDSGDSIEVSRNVEIVVKIISVFVDALPHVPEHRRPLPIIVQIVDTLGAEKFLWILLILL 60
QY 1420 PQYVTKTVLAAAYGK 1436
|||||
DB 61 PQYVTKTVLAAAYGK 77

RESULT 15
US-09-757-028-2590
Sequence 2590, Application US/09/57028
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PM001
CURRENT APPLICATION NUMBER: US/09/757,028
CURRENT FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
NUMBER OF SEQ ID NOS: 2660
SOFTWARE: Patent In Ver. 2.0

Query Match 3.7%; Score 79; DB 21; Length 164;
Best Local Similarity 100.0%; Pred No. 2e-67; 0; Indels 0; Gaps 0;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1519 SPMSQLSSNNFLKVVESGPEILKGLERLLFTVLGYISAVAQSMERNADKLTVKFWR 1578
|||||
Db 1 SPMSQLSSNNFLKVVESGPEILKGLERLLFTVLGYISAVAQSMERNADKLTVKFWR 60

QY 1579 AILSKAYDILLDKVNALLPT 1597
|||||
Db 61 AILSKAYDILLDKVNALLPT 79

RESULT 12
PCT-US01-006663-30151
; Sequence 30151, Application PC/TUS0100663
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Penn, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ANALYSIS OF GENE EXPRESSION IN HUMAN PLACENTA
; CURRENT APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04 02 00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 27 May 2000 (26 05 00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03-08-00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27 09 00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21-09-00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38837
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 30151
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP To A*G10807 2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EST HUMAN HIT: AL040338.1, EVALUE 8.00e-37
; OTHER INFORMATION: SWISSPROT HIT: P42945, EVALUE 2.00e-03
PCT-US01-006663-30151

Query Match 3.6%; Score 77; DB 1; Length 77;
Best Local Similarity 100.0%; Pred No. 8.9e-66;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SDSGDSIEVSRNVEIVVKIISVFDALPHVPEHRLPILVOLDTLGAERFLWILL 1419
|||||
Db 1 SDSGDSIEVSRNVEIVVKIISVFDALPHVPEHRLPILVOLDTLGAERFLWILL 60

QY 1420 PQYVTKTVLAAAYGEK 1436
|||||
Db 61 PQYVTKTVLAAAYGEK 77

RESULT 13
US-09-864-761-37036
; Sequence 37036, Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.

QY 1959 VRPFADTL 1966
|||||
Db 330 VKPFAUTL 337

RESULT 10
PCT-US01-18569-2412
; Sequence 2412, Application PC/TUS0118569
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P413PCT
; CURRENT APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2412
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (128)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-18569-2412

Query Match 5.9%; Score 126; DB 1; Length 135;
Best Local Similarity 100.0%; Pred No. 2.4e-113;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSLAQQQLKALPQSVASLLSRDEVASLLFDPKKAATIDRDTAFATCCTCLELLGIDP 60
|||||
Db 2 MTSLAQQQLKALPQSVASLLSRDEVASLLFDPKKAATIDRDTAFATCCTCLELLGIDP 61

QY 61 SFQGFAPFLSQLAKTLERSVQIKAVNKQIDENISLELIHLSPYFLKPAQKCLEWLIHR 120
|||||
Db 62 SFQGFAPFLSQLAKTLERSVQIKAVNKQIDENISLELIHLSPYFLKPAQKCLEWLIHR 121

QY 121 FHIHLY 126
|||||
Db 122 FHIHLY 127

RESULT 11
US-09 758 449-1058
; Sequence 1058, Application US/09758449
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM026
; CURRENT APPLICATION NUMBER: US/09/758,449
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1478
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1058
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-449-1058

14b	90	AEVTSTLEALATPQIPSLMPNLIITMKNTSELVSNFYVLSAIANLQKVETLPHFLSPY	149
177a	177a	LESTLSQVTHIEKLTSEMGSSAQNIRLTSLKKTATLTLAPRVLLPAIKKYQITKNNWK	1838
180b	150	LESTLSQVTHIEKLTSEMGSSAQNIRLTSLKKTATLTLAPRVLLPAIKKYQITKNNWK	209
183a	183a	NIMGPMSTLQIEHIGXMKKEELTSHQSQTAPFLTALDFAQHSSENDLEEGWGTENCILID	1898
180b	210	NIMGPMSTLQIEHIGXMKKEELTSHQSQTAPFLTALDFAQHSSENDLEEGWGTENCILID	269
189a	189a	CLVAMVVKLSFVTFPPFFKLTFWAKTFLADPDELLTFYNNLATGIAEKLKGLTTLFAGHL	1958
180b	270	CLVAMVVKLSFVTFPPFFKLTFWAKTFLADPDELLTFYNNLATGIAEKLKGLTTLFAGHL	359
195a	195a	VKFEADTL	1966
180b	330	VKFEADTL	337

```

1  RESULT      8
2  US 60 042 611 52
3  : Sequence 52, Application US/60042611
4  : GENERAL INFORMATION:
5  : APPLICANT: Bowcock, Anne M.
6  : APPLICANT: Bart, Richard
7  : TITLE OF INVENTION: Compositions and Methods Comprising
8  : TITLE OF INVENTION: BARD1 and other BRCA1 Binding Proteins
9  : NUMBER OF SEQUENCES: 115
10 : CORRESPONDENCE ADDRESS:
11 : ADDRESS: Arnold, White & Durkee
12 : STREET: P.O. Box 4539
13 : CITY: Houston
14 : STATE: Texas
15 : COUNTRY: USA
16 : ZIP: 77210
17 : COMPUTER READABLE FORM:
18 : MEDIUM TYPE: Floppy disk
19 : COMPUTER: IBM PC compatible
20 : OPERATING SYSTEM: PC-DOS/MS-DOS
21 : SOFTWARE: Patent In Release
22 : CURRENT APPLICATION DATA:
23 : APPLICATION NUMBER: US/60/042,611
24 : FILING DATE:
25 : CLASSIFICATION:
26 : ATTORNEY/AGENT INFORMATION:
27 : NAME: Hibler, David W.
28 : REGISTRATION NUMBER: P 41,071
29 : REFERENCE NUMBER: 499F22
30 : TELECOMMUNICATION INFORMATION:
31 : TELEPHONE: 512/418-4000
32 : TELEFAX: 512/474-7577
33 : INFORMATION FOR SEQ ID NO: 52:
34 : SEQUENCE CHARACTERISTICS:
35 : LENGTH: 515 amino acids
36 : TYPE: amino acid
37 : STRANDEDNESS:
38 : TOPOLOGY: linear
39 :
40 : US 60 042 611 52

```

QY	1719	A E V T S T L F A I A I P Q I P S L M P S L I T T M K N T S E L V S E V Y L L S A A A L G Q V V E T L P H T I S P Y	1778
Dd	90	A E V T S T L F A I A I P Q I P S L M P S L I T T M K N T S E L V S E V Y L L S A A A L G Q V V E T L P H T I S P Y	149
QY	1774	L E C H I S Q V I L R E K I T S E M G S A Q N R I T S L K K I T A T T L A P P V I L P A I K K T Y K G I E K N W K	1838
Dd	150	L E V I L S Q V I L E K I T S E M G S A Q N R I T S L K K I T A T T L A P P V I L P A I K K T Y K G I E K N W K	209

```

QY 1839 NHMGPPMSILQEHIGXMKKEELISHSQSQTAFPTLEALDFRAJHSENDEEVGKTENIID 1899
Db 210 NHMGPPMSILQEHIGXMKKEELISHSQSQTAFPTLEALDFRAJHSENDEEVGKTENIID 269
QY 1899 CIAMVVKI SEVTFPLPKIFDWAKTEADPKDILLTFYNIADICIAKLIKGLFTLFACHL 1958
Db 270 CIAMVVKI SEVTFPLPKIFDWAKTEADPKDILLTFYNIADICIAKLIKGLFTLFACHL 329
QY 1959 VKPFADTL 1666
Db 330 VKPFADTL 337

RESULT 9
US-60-042-985-52
: Sequence 52, Application US/60042985
: GENERAL INFORMATION:
: APPLICANT: Bowcock, Anne M.
: APPLICANT: Baer, Richard
: TITLE OF INVENTION: Compositions and Methods Comprising
: TITLE OF INVENTION: HARD and Other HCCA Binding Proteins
: NUMBER OF SEQUENCES: 115
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.40
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/30/042,985
: FILING DATE: Concurrently Herewith
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hibler, David W.
: REGISTRATION NUMBER: P-41,071
: REFERENCE/DOCKET NUMBER: USDS:499PZ3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 512/418-3000
: TELEFAX: 512/474-7577
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 515 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: US-60-042-985-52

```

[illegible]

QY 241 KLF 243
DB 242 KLF 244

RESULT 6
US-60-339-453-167
: Sequence 167, Application US/60339453
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Wang, Zhiwei
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Ghosh, Malabika
: APPLICANT: Weng, Gezhi
: APPLICANT: Boyle, Bryan J
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: POLYPEPTIDES
: FILE REFERENCE: 812
: CURRENT APPLICATION NUMBER: US/60/339,453
: PRIOR FILING DATE: 2002-03-12
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: PRIOR FILING DATE: 2000-02-28
: PRIOR APPLICATION NUMBER: US 09/577,409
: PRIOR FILING DATE: 2000-05-18
: PRIOR APPLICATION NUMBER: PCT/US01/04927
: PRIOR FILING DATE: 2001-02-26
: PRIOR APPLICATION NUMBER: US 09/519,705
: PRIOR FILING DATE: 2000-03-07
: PRIOR APPLICATION NUMBER: US 09/574,454
: PRIOR FILING DATE: 2000-05-19
: PRIOR APPLICATION NUMBER: PCT/US01/04941
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: US 09/643,167
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US01/08631
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: US 09/552,929
: PRIOR FILING DATE: 2000-04-18
: PRIOR APPLICATION NUMBER: US 09/770,160
: PRIOR FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: PCT/US01/08656
: PRIOR FILING DATE: 2001-04-18
: NUMBER OF SEQ ID NOS: 412
: SOFTWARE: pt_Fligenes Version 6.0
: SEQ ID NO 167
: LENGTH: 1149
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-339-453-167

Query Match 10.3%, Score 221, 5B 26, Length 1149;
Best Local Similarity 100.0%, Pred. No. 3.9e-205;
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1258 CLLNICQKLSFGGKTPKQILDEEKNVELIVQICIRLSEMPQTHHHALLLLGTIVAGIFPD 1317
DB 413 CLLNICQKLSFGGKTPKQILDEEKNVELIVQICIRLSEMPQTHHHALLLLGTIVAGIFPD 472
QY 1318 KVLHNIMSIETFGANVMRLDDTYSFOVINKTVKMWIPALIQSDSGDSIEVSRNVEEIV 1377
DB 473 KVLHNIMSIETFGANVMRLDDTYSFOVINKTVKMWIPALIQSDSGDSIEVSRNVEEIV 532
QY 1378 KIISVFDALPHVPEHRRKLPILVOLVDITGARKPFIWLLILLPQYVTKIVIAAAYGKGD 1437
DB 533 KIISVFDALPHVPEHRRKLPILVOLVDITGARKPFIWLLILLPQYVTKIVIAAAYGKGD 592
QY 1438 ALLEADTEFWFSVCCFESVGHQUSLNNILQYLKLPKEKE 1478
DB 593 ALLEADTEFWFSVCCFESVGHQUSLNNILQYLKLPKEKE 633

RESULT 7
US-08-936-487-52
: Sequence 52, Application US/08936487
: GENERAL INFORMATION:
: APPLICANT: Howcock, Anne M.
: APPLICANT: Baer, Richard
: TITLE OF INVENTION: Compositions and Methods Comprising
: TITLE OF INVENTION: BARD1 and other BRCA1 Binding Proteins
: NUMBER OF SEQUENCES: 130
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/936,487
: FILING DATE: Concurrently Herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/025,296
: FILING DATE: 20-SEP-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/042,611
: FILING DATE: 03-APR-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/042,985
: FILING DATE: 04-APR-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Hibler, David W.
: REGISTRATION NUMBER: P-41,071
: REFERENCE/DOCKET NUMBER: ITSD-499
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 512/418-3000
: TELEFAX: 512/474-7577
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 515 amino acids
: TYPE: amino acid
: STRANDNESS:
: TOPOLOGY: linear
US-08-936-487-52

Query Match 6.9%, Score 147, DB 13, Length 515;
Best Local Similarity 99.6%, Pred. No. 2.8e-133;
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1719 AEVSTLEALAIPLPSLMLPSLLTTMKNTSEIVSSEVYLLSALAALQKVETLPIHIFISPY 1778

```
QY 808 KSDIWNPEQLKESQVYHILITGLFEMMIKCAVAVHIEFVLMKFLFKVILLEVDFQLKFC 867
Db 722 KSDIWNPEQLKESQVYHILITGLFEMMIKCAVAVHIEFVLMKFLFKVILLEVDFQLKFC 781
QY 808 SVLWYTGSSLSNPINQVKTIVLQALVGGATVGGTAMISSKLTQKHQIASISSVAVVSELLNL 927
Db 782 SVLWYTGSSLSNPINQVKTIVLQALVGGATVGGTAMISLSSQKLTQKHQIASISSPVVSELLNL 941
QY 928 QTVKVEKRAALQVLAALSSVAGHFFYLLIDHETSKABETLSAAAVVIGGATLPLKEELKE 987
Db 842 QTVKVEKRAALQVLAALSSVAGHFFYLLIDHETSKABETLSAAAVVIGGATLPLKEELKE 901
QY 988 KKLKSHQKLSSTLKNLSGVYTSYPTAKDLMKVQGVVYKEMVLSQLIPMAEQLEKTK 1047
Db 902 KKLKSHQKLSSTLKNLSGVYTSYPTAKDLMKVQGVVYKEMVLSQLIPMAEQLEKTK 961
QY 1048 EPTAVIKDFAMVILHILITCKYNEFSVSLNEDPKSLDITKAVHILKELYAGMPTLIQIAL 1107
Db 962 EPTAVIKDFAMVILHILITCKYNEFSVSLNEDPKSLDITKAVHILKELYAGMPTLIQIAL 1021
QY 1108 EKITKDFPAAISURKVOOKLUMIFDLVNCNKNCHCAUTVSSVFKGTSVNAPOVRIEELP 1167
Db 1022 EKITKDFPAAISURKVOOKLUMIFDLVNCNKNCHCAUTVSSVFKGTSVNAPOVRIEELP 1081
QY 1168 PDKAKPLATVQKKRQKQKKSQDLESQVGGSSVWQPVTLILLELQHKKPLSPQILV 1227
Db 1082 PDKAKPLATVQKKRQKQKKSQDLESQVGGSSVWQPVTLILLELQHKKPLSPQILV 1141
QY 1228 PTFENLSKQLEPTEPOHOCNMXYTKOLLSLNTICOKLSHGCKIPKHILDEKFNKVL 1287
Db 1142 PTFENLSKQLEPTEPOHOCNMXYTKOLLSLNTICOKLSHGCKIPKHILDEKFNKVL 1201
QY 1288 IVQCTRLSEMPQTHRHALLILITVAGIFPDVKVLNIMSIPTMCGANVWHLDITYSFOVIN 1347
Db 1202 IVQCTRLSEMPQTHRHALLILITVAGIFPDVKVLNIMSIPTMCGANVWHLDITYSFOVIN 1261
QY 1348 KIVKAVTPALITDSQSDSTFVSNNVEEIVWKEIISVEVALPHVPEIRPLPLVQLVDTLG 1407
Db 1262 KIVKAVTPALITDSQSDSTFVSNNVEEIVWKEIISVEVALPHVPEIRPLPLVQLVDTLG 1321
QY 1408 AKKPIWILLILITFOYVTKIVLAAAYGRADILADLIEFWISVCEFSVQHOIQLSMNLI 1467
Db 1322 AKKPIWILLILITFOYVTKIVLAAAYGRADILADLIEFWISVCEFSVQHOIQLSMNLI 1381
QY 1468 QVILKIPKPKFTTHKAVSPKSSQREMLAVENVEIHLISQLEHFKPLSVSFMSCILSS 1527
Db 1382 QVILKIPKPKFTTHKAVSPKSSQREMLAVENVEIHLISQLEHFKPLSVSFMSCILSS 1441
QY 1528 NNEFKKVVESGPFILKGLLEFLLITVLYTISAVAGSMERNADKLTIVRTWRALLSKAYDL 1587
Db 1442 NNEFKKVVESGPFILKGLLEFLLITVLYTISAVAGSMERNADKLTIVRTWRALLSKAYDL 1501
QY 1588 LQKYNALLPTEIFDVIVKCLVGNPLPSVRRKALDLINNKLOONISWKKLTIVRTFKIAYDL 1647
Db 1502 LQKYNALLPTEIFDVIVKCLVGNPLPSVRRKALDLINNKLOONISWKKLTIVRTFKIAYDL 1561
QY 1648 LLAIVORK 1655
Db 1562 LLAIVORK 1569
```

RESULT 5

US-60-339-453 458

Sequence 158, Application US/60339453

GENERAL INFORMATION:

APPLICANT: Tana, Y. Tom

APPLICANT: Wang, Zhiwei

APPLICANT: Wang, Jian-Rui

APPLICANT: Ghosh, Malabika

APPLICANT: Wong, George

APPLICANT: Boyle, Bryan J

APPLICANT: Dmanan, Rajaraj T

TITLE OF INVENTION: Novel Nucleic Acids and

```

: TITLE OF INVENTION: Polypeptides
: FILE REFERENCE: 812
: CURRENT APPLICATION NUMBER: US/60/339,453
: CURRENT FILING DATE: 2002-03-12
: PRIOR APPLICATION NUMBER: US/99/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US/99/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US/99/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US/99/499,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US/99/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03900
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US/99/515,126
: PRIOR FILING DATE: 2000-02-23
: PRIOR APPLICATION NUMBER: US/99/577,409
: PRIOR FILING DATE: 2000-05-13
: PRIOR APPLICATION NUMBER: PCT/US01/04927
: PRIOR FILING DATE: 2001-02-25
: PRIOR APPLICATION NUMBER: US/99/519,705
: PRIOR FILING DATE: 2000-03-07
: PRIOR APPLICATION NUMBER: US/99/574,454
: PRIOR FILING DATE: 2000-05-13
: PRIOR APPLICATION NUMBER: PCT/US01/04941
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US/99/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: US/99/649,167
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US01/08641
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: US/99/552,929
: PRIOR FILING DATE: 2000-04-18
: PRIOR APPLICATION NUMBER: US/99/770,160
: PRIOR FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: PCT/US01/08656
: PRIOR FILING DATE: 2001-04-18
: NUMBER OF SEQ ID NOS: 412
: SOFTWARE: PTFigures Version 6.0
: SEQ ID NO 458
: LENGTH: 325
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-60-339-453-358
```

Query Match 11.3%, Score 243; DB 26; Length 325;

Best Local Similarity 100.0%; Pred. No. 4,1e 237;

Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MTSIAQQLQRIALPOSDASLLSKDFVVASLLFDPKAAATIDRDTAFALGCTIGLELLGIDP 66
Db 2 MTSIAQQLQRIALPOSDASLLSKDFVVASLLFDPKAAATIDRDTAFALGCTIGLELLGIDP 61
QY 61 SFEQFEAPLFSQIAKTLEPSVATKAVNKGIDFNTISLFIHLSPYFLKPAQKCLEWLIHR 120
Db 62 SFEQFEAPLFSQIAKTLEPSVATKAVNKGIDFNTISLFIHLSPYFLKPAQKCLEWLIHR 121
QY 121 FHIHLYNODSLIACVLPYHETIEFVRVIGLLKINNSKHEFWLLPVKQSGVPLAKGTLTIT 180
Db 122 FHIHLYNODSLIACVLPYHETIEFVRVIGLLKINNSKHEFWLLPVKQSGVPLAKGTLTIT 181
QY 181 HCYKDLGDMDFICSILVIVKSVKFAHYPPSSAQIKVLLAFYASIIISALVAADVSNIITA 240
Db 182 HCYKDLGDMDFICSILVIVKSVKFAHYPPSSAQIKVLLAFYASIIISALVAADVSNIITA 241
```

301 LSCILVLLQORQESIGKPPFPCNVPDLITILHIGISSETYDVSPLLRYMLPHLVVSIH 360
QY
361 HVIGRFTEDMAGQYKRRHLEAIIKISIKNNIDHLIASILLFEYISYSSQEMDSNKSVS 420
DB
361 HVIGRFTEDMAGQYKRRHLEAIIKISIKNNIDHLIASILLFEYISYSSQEMDSNKSVS 420
QY
421 LNEOFPLIRLIESKYPRTLDDVVLPEHLKEIADLKQOELFHOFVSTSTGGKYQFLADSD 480
DB
421 LNEOFPLIRLIESKYPRTLDDVVLPEHLKEIADLKQOELFHOFVSTSTGGKYQFLADSD 480
QY
481 TSIMLSLNPLAPVRILAMNHLKIMKTSKEGVDSFTKEPAVILARIQDINDIVLSAISA 540
DB
481 TSIMLSLNPLAPVRILAMNHLKIMKTSKEGVDSFTKEPAVILARIQDINDIVLSAISA 540
QY
541 PEIPEKHFSESVTISNLIINIPORAKLSKNGWYEVILKIAADILIKKEIISENIQLSNQV 600
DB
541 PEIPEKHFSESVTISNLIINIPORAKLSKNGWYEVILKIAADILIKKEIISENIQLSNQV 600
QY
601 VCLLPFWINDDTESAEMKIAIYLSKSGICSLHPLRGWEEALENVIKSTKPKLIGVA 660
DB
601 VCLLPFWINDDTESAEMKIAIYLSKSGICSLHPLRGWEEALENVIKSTKPKLIGVA 660
QY
661 NQKMIELIAJININIGUPSSMKMVEDLISVGRHESFNIAKOKVTFFHVLISVLVSCCSSLIKE 720
DB
661 NQKMIELIAJININIGUPSSMKMVEDLISVGRHESFNIAKOKVTFFHVLISVLVSCCSSLIKE 720
QY
721 THPPPAIRVFSLLQKKIKKLESVITAVEIPSWHIFELMDRGIPVELWAHYVRELINSTOR 780
DB
721 THPPPAIRVFSLLQKKIKKLESVITAVEIPSWHIFELMDRGIPVELWAHYVRELINSTOR 780
QY
781 VAVEDSVLVSLLKFIYALKAPSKPKPTIWNPNQLKPKSPDYDILTLGLFEMMINGA 840
DB
781 VAVEDSVLVSLLKFIYALKAPSKPKPTIWNPNQLKPKSPDYDILTLGLFEMMINGA 840
QY
841 DAYHFPVIMKFLPKVHLDPVLPFKCSVLWMTYCSSINPLNCSVKITVLTQALYVCCAM 900
DB
841 DAYHFPVIMKFLPKVHLDPVLPFKCSVLWMTYCSSINPLNCSVKITVLTQALYVCCAM 900
QY
901 LSSQKTOCKHOTIASISSPVVTSILINIGSPVKEVRKAAIUCIQAISVASPPYLIIDHLI 960
DB
901 LSSQKTOCKHOTIASISSPVVTSILINIGSPVKEVRKAAIUCIQAISVASPPYLIIDHLI 960
QY
961 SKAEEITSDAAYVILQIALIFELQREKKLKSQKSETIKNLLSCVYSCPSYIAKDLMK 1020
DB
961 SKAEEITSDAAYVILQIALIFELQREKKLKSQKSETIKNLLSCVYSCPSYIAKDLMK 1020
QY
1021 VLOCVNEMLVLSQILMAEQILEKTOKEPTIAVLKDPAMVILHETLGKYNFESVSLINEDPK 1080
DB
1021 VLOCVNEMLVLSQILMAEQILEKTOKEPTIAVLKDPAMVILHETLGKYNFESVSLINEDPK 1080
QY
1081 SLDIFIKAVHTIKELIYACMPITQITALEKITKPPFAAISDEKVOOKILIRMIIDILVNCNK 1140
DB
1081 SLDIFIKAVHTIKELIYACMPITQITALEKITKPPFAAISDEKVOOKILIRMIIDILVNCNK 1140
QY
1141 SHCAQTVSSVFEKGISVNAEGVRIEPPDFAKPLGTGQKPPAKMUKKSGLESVUEV 1200
DB
1141 SHCAQTVSSVFEKGISVNAEGVRIEPPDFAKPLGTGQKPPAKMUKKSGLESVUEV 1200
QY
1201 GSYWQVRLILELLOHKKKURSFGVILVPLFNLSICELPIPOGCGNMYTKQLILSCLL 1260
DB
1201 GSYWQVRLILELLOHKKKURSFGVILVPLFNLSICELPIPOGCGNMYTKQLILSCLL 1260
QY
1261 NICOKLSIDGCKIPKIDIDREKFNVELIVQIPILSEMHQTHHALLILGIVAGIFPDKVL 1320
DB
1261 NICOKLSIDGCKIPKIDIDREKFNVELIVQIPILSEMHQTHHALLILGIVAGIFPDKVL 1320
QY
1321 IHTMSIFTEMCANVMPLDDTYSFQVINKTVKMWIPALIQSDSGDSIEVSPNVEEIVVKII 1380
DB
1321 IHTMSIFTEMCANVMPLDDTYSFQVINKTVKMWIPALIQSDSGDSIEVSPNVEEIVVKII 1380
QY
1381 SVFVDALPHVPEHRRPILVQIVDTHCAEKPLIWLIIILFQYVTKVLAAYGCKDAIL 1440
DB
1381 SVFVDALPHVPEHRRPILVQIVDTHCAEKPLIWLIIILFQYVTKVLAAYGCKDAIL 1440

1441 EADTEWFSWCCEFSYOHQIQSLMNILOYLKLUPKEKEETIPKAVSFNKSSESQEMLOVF 1500
DB
1441 EADTEWFSWCCEFSYOHQIQSLMNILOYLKLUPKEKEETIPKAVSFNKSSESQEMLOVF 1500
QY
1501 NVETHTSKOLRIHFKFLSVSFMSSQLSSNNFLKKVSVESGPEILKGLERLETVLGYISA 1560
DB
1501 NVETHTSKOLRIHFKFLSVSFMSSQLSSNNFLKKVSVESGPEILKGLERLETVLGYISA 1560
QY
1561 VAGSMERNADKLTVKWRALLSKAYDLDDKVNALLPTETFPVIRGIVCNPILPSVRKAL 1620
DB
1561 VAGSMERNADKLTVKWRALLSKAYDLDDKVNALLPTETFPVIRGIVCNPILPSVRKAL 1620
QY
1621 DLLNNKLOQNISMKKTIVTPELKLVPDLAIVQKKKEGEEQAINPQTALYTIKILCKN 1680
DB
1621 DLLNNKLOQNISMKKTIVTPELKLVPDLAIVQKKKEGEEQAINPQTALYTIKILCKN 1680
QY
1681 FGAENPDPPFVVLIXTAVKLIAPERKEEKNVGLSALLCIAEVTSTLEALAIQPLPSLMPSL 1740
DB
1681 FGAENPDPPFVVLIXTAVKLIAPERKEEKNVGLSALLCIAEVTSTLEALAIQPLPSLMPSL 1740
QY
1741 LTTMKNTSELVSSPVYLLSALAALOKVVFLLPHFISPYLRGILISOVILHKKITSEMGAS 1800
DB
1741 LTTMKNTSELVSSPVYLLSALAALOKVVFLLPHFISPYLRGILISOVILHKKITSEMGAS 1800
QY
1801 QANIRLTSKKTILATLAPRVLLPAIKTYKQIEKNKNHMGPFMSILOEIHGXMKKEEL 1860
DB
1801 QANIRLTSKKTILATLAPRVLLPAIKTYKQIEKNKNHMGPFMSILOEIHGXMKKEEL 1860
QY
1861 TSISQSUTAFLEALUPKAHSENDLEEVCKIENCILICLVAMVVKLSEVTFRPLFKLF 1920
DB
1861 TSISQSUTAFLEALUPKAHSENDLEEVCKIENCILICLVAMVVKLSEVTFRPLFKLF 1920
QY
1921 DWAKTEDAPKDRILITFYNLADCIAEKLKGIFTIPACHLVKPPADTL 1966
DB
1921 DWAKTEDAPKDRILITFYNLADCIAEKLKGIFTIPACHLVKPPADTL 1966

RESULT 4
PCT-US01-08631-45591
; Sequence 45591, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 45591
; LENGTH: 15569
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: DOMAIN
; LOCATION: (416)..(427)
; OTHER INFORMATION: Caspains alpha/beta proteolysis domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL003069, p-value=7.300e-09, raw score of 8.2
PCT-US01-08631-45591

Query Match 42.4%; Score 908; DB 1; Length 1569;
Best Local Similarity 100.0%; Pred No. 0;
Matches 908; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 EIPSEWHIELMDRGIPVELWAHYVEELNSTQRAVEUSVFLVSLKFIYALKAPKSEFP 807
DB 662 EIPSEWHIELMDRGIPVELWAHYVEELNSTQRAVEUSVFLVSLKFIYALKAPKSEFP 721

```

1481 SVFVIALPHVPEHRLFLVLAVALVDTLAEKFLWIIIIIFEQYVTKVLAAAYREKDAIIL 1440
1441 EADTEFWFSVCFESVCHQJOSLMNLTQYLLKTPREKFTIPKAVSPKNSQSFQEMLVGF 1500
1441 EADTEFWFSVCFESVCHQJOSLMNLTQYLLKTPREKFTIPKAVSPKNSQSFQEMLVGF 1500
1501 NVETHTSKQIHPKFLSVSEMSQGLSSNNFLKVVESGSPPEILKGLPELLETVLGYISA 1560
1501 NVETHTSKQIHPKFLSVSEMSQGLSSNNFLKVVESGSPPEILKGLPELLETVLGYISA 1560
1561 VAGSWERNADWITVYKFRALLSKATVLLIKVNAIIPTFTFIPVITPVLVYNLPSVPPKAL 1620
1561 VAGSWERNADWITVYKFRALLSKATVLLIKVNAIIPTFTFIPVITPVLVYNLPSVPPKAL 1620
1621 DILLNKKLQUNISWKKIIVIKRKIVPPIIAIVORKKKKEDEPOAINROTALYILKILCKN 1680
1621 DILLNKKLQUNISWKKIIVIKRKIVPPIIAIVORKKKKEDEPOAINROTALYILKILCKN 1680
1681 EGAENPDFFVVI XTAVKI IAPERKFFNVLSALLT'IAEVTSTLEALATPQIPSDMPSE 1740
1681 EGAENPDFFVVI XTAVKI IAPERKFFNVLSALLT'IAEVTSTLEALATPQIPSDMPSE 1740
1741 LITMKNTSEIIVSEVYIISALAAKQVVEIIPHEIETSPYLEIISAVIHLEKTISEWSSAS 1800
1741 LITMKNTSEIIVSEVYIISALAAKQVVEIIPHEIETSPYLEIISAVIHLEKTISEWSSAS 1800
1801 QANIKLSKKTALIAIRVLPATKKYKQKTKNNKNNHMGPPNSIIQERIGAMKKKEEL 1860
1801 QANIKLSKKTALIAIRVLPATKKYKQKTKNNKNNHMGPPNSIIQERIGAMKKKEEL 1860
1861 TSHQSQTAFTHI EALDFPAQHSFNILEVFKTENIILP'LVAMVVKLSEVTPRPPLFFKIF 1920
1861 TSHQSQTAFTHI EALDFPAQHSFNILEVFKTENIILP'LVAMVVKLSEVTPRPPLFFKIF 1920
1921 DWAKTEDAKKRLITFYNLADQIAEKIKHIFTIFPAHLVKKPFAITLXAVNISKIDAEAFD 1980
1921 DWAKTEDAKKRLITFYNLADQIAEKIKHIFTIFPAHLVKKPFAITLXAVNISKIDAEAFD 1980
1981 SENDPKKCTLIQFUNGILYKIFEDIOHFIISKERAXALMMDI VQDLENRLOGEKKFOHR 2040
1981 SENDPKKCTLIQFUNGILYKIFEDIOHFIISKERAXALMMDI VQDLENRLOGEKKFOHR 2040
2041 VTKHLIPVIAQSVAMAGDSLWKPLNYGTLTKTRESSEKVFPAALITVLALAEKIKENYI 2100
2041 VTKHLIPVIAQSVAMAGDSLWKPLNYGTLTKTRESSEKVFPAALITVLALAEKIKENYI 2100
2101 VLLPESIPPLAEIMPEPFFVPHVQVFTTQAEFTVLGEPLQSYF 2144
2101 VLLPESIPPLAEIMPEPFFVPHVQVFTTQAEFTVLGEPLQSYF 2144

```

```

RESULT
US-60-141-424-3
Sequence 5, Application US/60141424
GENERAL INFORMATION:
APPLICANT: Barry, Caroline
APPLICANT: Bouqueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: A NOVEL BAP28 GENE AND PROTEIN
FILE REFERENCE: GENSET.063PRF
CURRENT APPLICATION NUMBER: US/60-141-424
CURRENT FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent .pm
SEQ ID NO: 3
LENGTH: 2144
TYPE: PRT
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 91..111
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program

```

```

FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 212..232
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 700..720
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 861..881
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 884..904
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 911..931
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 937..957
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 1297..1317
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 1410..1430
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 1725..1745
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 1987..2007
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: SITE
LOCATION: 1750..1771
OTHER INFORMATION: potential leucine zipper pattern, SLA Program
US-60-141-323-3

```

```

Query Match 82.3%, Score 1764, DB 24, Length 2144:
Best Local Similarity 99.9%, Pred. No. 0:
Matches 1964: Conservative 0; Mismatches 2; Indels 0; Gaps 0:

QY 1 MTSIAGQIQRIALPOSDASLSRDEVASLIFDVKAAIIDKDIAPALQVIGIEELIGIDP 60
DB 1 MTSIAGQIQRIALPOSDASLSRDEVASLIFDVKAAIIDKDIAPALQVIGIEELIGIDP 60
QY 61 SEQFEAPLPSQLAKTLERSVQTKAVNKGIDENISLFIHLSPYFLKPKACKGLEWLIHR 120
DB 61 SEQFEAPLPSQLAKTLERSVQTKAVNKGIDENISLFIHLSPYFLKPKACKGLEWLIHR 120
QY 121 FTHILYNQSLIA'VLPYHETPIFVPTVLI'KINN'PFPWFELPVY'GVPLAKFTLIT 180
DB 121 FTHILYNQSLIA'VLPYHETPIFVPTVLI'KINN'PFPWFELPVY'GVPLAKFTLIT 180
QY 181 HCYKDIGEMDFICSLVTKSVKVAEYDGSAAQIRVLIAPYASTIVSALVAAEIVSINIIA 240
DB 181 HCYKDIGEMDFICSLVTKSVKVAEYDGSAAQIRVLIAPYASTIVSALVAAEIVSINIIA 240
QY 241 KLFPPYTKGLSKSLPDYRAATYIMIQISVKVTMENTFVNSLASQITKLTIKTSPSLKDG 300
DB 241 KLFPPYTKGLSKSLPDYRAATYIMIQISVKVTMENTFVNSLASQITKLTIKTSPSLKDG 300
QY 301 LSTIVLILKQKPESEIGKKEFFHILNVPIQIILHILSIEIYVNSPILPFWIPIHVVSIH 360
DB 301 LSTIVLILKQKPESEIGKKEFFHILNVPIQIILHILSIEIYVNSPILPFWIPIHVVSIH 360

```

```

Db 1981 SENDPKCCCLLGLFLLNCLYK: FLEDTQHFISKERAXALMMPVLQGLNKLQGEKFKQER 2040
QY 2041 VTKLLPCLAQFSAVAMADDSLWKPFLNYGILLKTGSSPKVRFPAALITVIALAEKLEKNEYI 2100
Db 2041 VTKLLPCLAQFSAVAMADDSLWKPFLNYGILLKTGSSPKVRFPAALITVIALAEKLEKNEYI 2100
QY 2101 VLLPESIPFLAEMEDPCHEVREHQCCQKTKQOQLFTVIGEPHQSYF 2144
Db 2101 VLLPESIPFLAEMEDPCHEVREHQCCQKTKQOQLFTVIGEPHQSYF 2144

RESULT 2
US-60-176-880-4
: Sequence 4, Application US/60176880
: GENERAL INFORMATION:
: APPLICANT: Bouqueleret, Lydie
: APPLICANT: Choumakov, Ilya
: TITLE OF INVENTION: A NOVEL HAP28 GENE AND PROTEIN.
: FILE REFERENCE: 67.US2.PRO
: CURRENT APPLICATION NUMBER: US/60/176.880
: EARLIER FILING DATE: 2000-01-18
: EARLIER APPLICATION NUMBER: US 60/141.323
: EARLIER FILING DATE: 1999-06-25
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: Patent.pm
: SEQ ID NO 4
: LENGTH: 2144
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: 1694
: OTHER INFORMATION: Xaa-Ser or Asn
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: 1854
: OTHER INFORMATION: Xaa-Ala or Val
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: 1967
: OTHER INFORMATION: Xaa-Asp or Asn
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: 2017
: OTHER INFORMATION: Xaa-Gly or Glu
US-60-176-880-4

Query Match 99.88; Score 2140; DB 26; Length 2144;
Best Local Similarity 100.0%; P-adj. No 0;
Matches 2144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSIAQOLQRIALPOSVASILSRDEVASILFDPKFAATIDRDYAFAGCTGCLPELIGIDP 60
Db 1 MTSIAQOLQRIALPOSVASILSRDEVASILFDPKFAATIDRDYAFAGCTGCLPELIGIDP 60
QY 61 SFEQFEAPLESQAKTLERSVQTKAVNKQLDENISLFIHLSEYFLLKPAQKCLEWLIHR 120
Db 61 SFEQFEAPLESQAKTLERSVQTKAVNKQLDENISLFIHLSEYFLLKPAQKCLEWLIHR 120
QY 121 FPHILYNQDSLACVLPYHETRIFFVRVIOILLKINNSHRWFLLIPVKQSGVPLAKGTLLT 180
Db 121 FPHILYNQDSLACVLPYHETRIFFVRVIOILLKINNSHRWFLLIPVKQSGVPLAKGTLLT 180
QY 181 HCYKDJGFMDFICSLVTKSVKVFARFPCSSAQILVILAFYASTIVSALVAEDVDSDNIIA 240
Db 181 HCYKDJGFMDFICSLVTKSVKVFARFPCSSAQILVILAFYASTIVSALVAEDVDSDNIIA 240
QY 241 KLPFYTKQKLSLSPDYRAATYMTIQISVKVTMENTFVNSLASQITKTLTKIPSLIKDG 300
Db 241 KLPFYTKQKLSLSPDYRAATYMTIQISVKVTMENTFVNSLASQITKTLTKIPSLIKDG 300

```

```

QY 360 LSLQILVILQPKPKPSLAKKPPPHLQNVNVDLITLHGISFTVDVSPILRYMILPHLWVSLIH 360
Db 360 LSLQILVILQPKPKPSLAKKPPPHLQNVNVDLITLHGISFTVDVSPILRYMILPHLWVSLIH 360
QY 361 HVTGEETEGMDGOIYKRLHEALITKISLKNLHDLLASLLFEEYISYSSQEMDSNKSYS 420
Db 361 HVTGEETEGMDGOIYKRLHEALITKISLKNLHDLLASLLFEEYISYSSQEMDSNKSYS 420
QY 421 LNEQFLPLIELLESKYPTLDVLEELHKEIADLKKUELHFQVLSLTSQKYFLAELSL 480
Db 421 LNEQFLPLIELLESKYPTLDVLEELHKEIADLKKUELHFQVLSLTSQKYFLAELSL 480
QY 481 TSLMLSLNHPLAPVRIIAMNHLKKIMKTSKGVDESPITKFAVIAELGDDNDIVLSAISA 540
Db 481 TSLMLSLNHPLAPVRIIAMNHLKKIMKTSKGVDESPITKFAVIAELGDDNDIVLSAISA 540
QY 541 FEIPKEHFSSEVTISNLLNLFORAELSKNGEWEYVLKIAADILKEILSENUGLSNOVV 600
Db 541 FEIPKEHFSSEVTISNLLNLFORAELSKNGEWEYVLKIAADILKEILSENUGLSNOVV 600
QY 601 VCTLPFVVIINDDTESAEMKTAIYLSKSGICSLHPLLPGRFEALENVIKSTKPKKLIGVA 660
Db 601 VCTLPFVVIINDDTESAEMKTAIYLSKSGICSLHPLLPGRFEALENVIKSTKPKKLIGVA 660
QY 661 NQKMIPELLADNINIGDPSSMKMVEDLISVGEESFNLKOKVTFHVILSVLSVSCSSILKE 720
Db 661 NQKMIPELLADNINIGDPSSMKMVEDLISVGEESFNLKOKVTFHVILSVLSVSCSSILKE 720
QY 721 THPPFAIRVFSLLQKKIKKLESVITAVEISEWHIELMDRGIPVELWVHYVEELNSTOR 780
Db 721 THPPFAIRVFSLLQKKIKKLESVITAVEISEWHIELMDRGIPVELWVHYVEELNSTOR 780
QY 781 VAVEDSVFVSLKKFIYALKAPKSPKGDWWNPQOLKEDSRQYLHLLIGLFEMMLNGA 840
Db 781 VAVEDSVFVSLKKFIYALKAPKSPKGDWWNPQOLKEDSRQYLHLLIGLFEMMLNGA 840
QY 841 DAVHFRVIMKLIPTKVHLEDVHQJFKPCSVIWTYSSISNPLNCVSKVIVQTAALVGCAM 900
Db 841 DAVHFRVIMKLIPTKVHLEDVHQJFKPCSVIWTYSSISNPLNCVSKVIVQTAALVGCAM 900
QY 901 ISSQKTKQKHQASISSPVVTSILINIGSPKVEPRPAALITVIALSVASPFYLLIDHLL 960
Db 901 ISSQKTKQKHQASISSPVVTSILINIGSPKVEPRPAALITVIALSVASPFYLLIDHLL 960
QY 961 SKAEIITSDAAVYIQGLATLFEELQREKLLKSHQKLSLKNLSSVYSPSYIAKDLMK 1020
Db 961 SKAEIITSDAAVYIQGLATLFEELQREKLLKSHQKLSLKNLSSVYSPSYIAKDLMK 1020
QY 1021 VLOQVNGEMVLSQILLPMARQLLEKIQKEPTAVIKDPAWVILHITICKYNEFSVSLINEDPK 1080
Db 1021 VLOQVNGEMVLSQILLPMARQLLEKIQKEPTAVIKDPAWVILHITICKYNEFSVSLINEDPK 1080
QY 1081 SLQIFIKAVHTTKELYAGMPTIQITALEKTKPPTFAAISDEKVOOKLIPMLFDIIVNCKN 1140
Db 1081 SLQIFIKAVHTTKELYAGMPTIQITALEKTKPPTFAAISDEKVOOKLIPMLFDIIVNCKN 1140
QY 1141 SHCAQTSSVSVFKGISVNAEQVPIELEPPKAKPIATVQKPPKAKMOOKKSDLESVAEVS 1200
Db 1141 SHCAQTSSVSVFKGISVNAEQVPIELEPPKAKPIATVQKPPKAKMOOKKSDLESVAEVS 1200
QY 1201 GSYWQKVTLLLELLOHKKILRSPQILVPLFNLISACGLEPILPQKQNMETTKQILISCL 1260
Db 1201 GSYWQKVTLLLELLOHKKILRSPQILVPLFNLISACGLEPILPQKQNMETTKQILISCL 1260
QY 1261 NICUKLSFGSGKIFKDIIDEEKFNVELIVQICIRLSEMPQTHHALLLLGTVAGIFPKVL 1320
Db 1261 NICUKLSFGSGKIFKDIIDEEKFNVELIVQICIRLSEMPQTHHALLLLGTVAGIFPKVL 1320
QY 1321 HNIMSIFTEGANYMRLDDTYSFOVINKTVKMVIPALIQSDSGSDSIEVSRNVEEIVVKII 1380
Db 1321 HNIMSIFTEGANYMRLDDTYSFOVINKTVKMVIPALIQSDSGSDSIEVSRNVEEIVVKII 1380
QY 1381 SVFVDAIPVHVEKHKRKLPIVQGLVVIHGAKKLHWLILILILFQGVYVIRVIAAAYGKDAIIL 1440

```

LOCATION: 1967									
OTHER INFORMATION: Xaa-Asp or Asn									
NAME/KEY: VARIANT									
LOCATION: 2017									
OTHER INFORMATION: Xaa-Gly or Glu									
US 09 603 665 5									
Query Match									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 2144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MISLAQULQRIALPOSDASLSRQVVASLLPDKKAATIDRDITAFAGCTGIFELIGDIP	60						
DB	1	MISLAQULQRIALPOSDASLSRQVVASLLPDKKAATIDRDITAFAGCTGIFELIGDIP	60						
QY	61	SEQFEAPLFSQAKTLETSVOTKAVNKULDENISLFLIHLSPYELLKPAQKLEWLIHR	120						
DB	61	SEQFEAPLFSQAKTLETSVOTKAVNKULDENISLFLIHLSPYELLKPAQKLEWLIHR	120						
QY	121	FHILHYNDUSLIACVLPYHETIRFVRVIGQLLKINNSKHFWLWLPKQSGVPLAKGTLLT	180						
DB	121	FHILHYNDUSLIACVLPYHETIRFVRVIGQLLKINNSKHFWLWLPKQSGVPLAKGTLLT	180						
QY	181	HYKYLQWDFICSLVKSQVFAEYPSAQLRVLLAFYASTIVSAIVAAEDVSNNIIA	240						
DB	181	HYKYLQWDFICSLVKSQVFAEYPSAQLRVLLAFYASTIVSAIVAAEDVSNNIIA	240						
QY	241	KLPFYIQGLKSSLPDYRAATYMIICQISVKYTMENFVNSLASQIIKTLTKIPSLIKDG	300						
DB	241	KLPFYIQGLKSSLPDYRAATYMIICQISVKYTMENFVNSLASQIIKTLTKIPSLIKDG	300						
QY	401	LSCVLIVLQRPESLGGKPPPHLNVDPDITILHGISETYDVSPLLRYMLPHLVVYSIHH	360						
DB	401	LSCVLIVLQRPESLGGKPPPHLNVDPDITILHGISETYDVSPLLRYMLPHLVVYSIHH	360						
QY	361	HVGFETEMQIQYKRRHFAILKLSKNNLHLLASHLFEYISYSSQEDMSNKVSL	420						
DB	361	HVGFETEMQIQYKRRHFAILKLSKNNLHLLASHLFEYISYSSQEDMSNKVSL	420						
QY	421	INEQFLPLIKLESYPTLDVLEHLEKETAQKKQELFHQVSLSTSGKYQFLADSD	480						
DB	421	INEQFLPLIKLESYPTLDVLEHLEKETAQKKQELFHQVSLSTSGKYQFLADSD	480						
QY	481	TSLSLSNIHPLAFVRIILAMNHLKKIMKTSKEGVDESFKKAVLARLGDENIDVLSAIS	540						
DB	481	TSLSLSNIHPLAFVRIILAMNHLKKIMKTSKEGVDESFKKAVLARLGDENIDVLSAIS	540						
QY	541	FEIPREHSSVITISNLNLFORAEISKNQWYEVLIKIAAGLILKEELISENDQLSNQVV	600						
DB	541	FEIPREHSSVITISNLNLFORAEISKNQWYEVLIKIAAGLILKEELISENDQLSNQVV	600						
QY	601	VLLFQVVIINNDTESAKMKIATYLSKSIITSLHPLLQWEEALENVKSTKPKGLIGVA	660						
DB	601	VLLFQVVIINNDTESAKMKIATYLSKSIITSLHPLLQWEEALENVKSTKPKGLIGVA	660						
QY	661	NKMKIFELADNINLGDPSMMIKWFFMLTSVFEESFNLCQKVTPIVILSVLYSGSSSLKE	720						
DB	661	NKMKIFELADNINLGDPSMMIKWFFMLTSVFEESFNLCQKVTPIVILSVLYSGSSSLKE	720						
QY	721	THEPFAIRVESLQKKIKKILLESVITAVEIPSPWHIEMIDRGIPWELWAHYVEELNSTGR	780						
DB	721	THEPFAIRVESLQKKIKKILLESVITAVEIPSPWHIEMIDRGIPWELWAHYVEELNSTGR	780						
QY	781	VAVEUSVFLVPSLKRFIYALKAPSPKIDWNNPEQLKEDSRDYLHLLIGLFEMMLNCA	840						
DB	781	VAVEUSVFLVPSLKRFIYALKAPSPKIDWNNPEQLKEDSRDYLHLLIGLFEMMLNCA	840						
QY	841	DAVHEFVLMKLFIKVHLEDFVQLEKFCVSWLWYSSLSNPLNCVSKVTLOTQALVYGCA	900						
DB	841	DAVHEFVLMKLFIKVHLEDFVQLEKFCVSWLWYSSLSNPLNCVSKVTLOTQALVYGCA	900						
QY	901	ISSQAQCNKHOIASISSPVVLSLLINDGSPVAVRRAAIQCIQALSGVASPPYLLIDHLL	960						

GenCore version 4.5
Copyright (c) 1993 2000 CompuGen Ltd

OM protein protein search, using sw model

Run on: July 2, 2002, 16:14:43 : Search time 154.98 Seconds
(without alignments)
4728.953 Million cell updates/sec

Title: US-09-603-665-5

Perfect score: 2144

Sequence: 1 MTSAGGQGFVIGSPVLSVF 2144

Scoring table:

Gapop 60 0 0 Gapext 60 0 0

Searched: 3592263 seqs, 3592263 residues

Word size: 6

Total number of hits satisfying chosen parameters: 34338

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Pending Patents, AA, Main: *

```

1: /cgn2_6/ptodata/2/paa/us086_COMB pep: *
2: /cgn2_6/ptodata/2/paa/us087_COMB pep: *
3: /cgn2_6/ptodata/2/paa/us088_COMB pep: *
4: /cgn2_6/ptodata/2/paa/us089_COMB pep: *
5: /cgn2_6/ptodata/2/paa/us090_COMB pep: *
6: /cgn2_6/ptodata/2/paa/us091_COMB pep: *
7: /cgn2_6/ptodata/2/paa/us092_COMB pep: *
8: /cgn2_6/ptodata/2/paa/us093_COMB pep: *
9: /cgn2_6/ptodata/2/paa/us094_COMB pep: *
10: /cgn2_6/ptodata/2/paa/us095_COMB pep: *
11: /cgn2_6/ptodata/2/paa/us096_COMB pep: *
12: /cgn2_6/ptodata/2/paa/us097_COMB pep: *
13: /cgn2_6/ptodata/2/paa/us098_COMB pep: *
14: /cgn2_6/ptodata/2/paa/us099_COMB pep: *
15: /cgn2_6/ptodata/2/paa/us100_COMB pep: *
16: /cgn2_6/ptodata/2/paa/us101_COMB pep: *
17: /cgn2_6/ptodata/2/paa/us102_COMB pep: *
18: /cgn2_6/ptodata/2/paa/us103_COMB pep: *
19: /cgn2_6/ptodata/2/paa/us104_COMB pep: *
20: /cgn2_6/ptodata/2/paa/us105_COMB pep: *
21: /cgn2_6/ptodata/2/paa/us106_COMB pep: *
22: /cgn2_6/ptodata/2/paa/us107_COMB pep: *
23: /cgn2_6/ptodata/2/paa/us108_COMB pep: *
24: /cgn2_6/ptodata/2/paa/us109_COMB pep: *
25: /cgn2_6/ptodata/2/paa/us110_COMB pep: *
26: /cgn2_6/ptodata/2/paa/us111_COMB pep: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2140	99.8	2144	20	US-09-603-665-5
2	2140	99.8	2144	26	US-09-603-665-5
3	1764	82.3	2144	26	US-09-603-665-5
4	908	42.4	1569	1	PCT-US01-08631-45591
5	243	11.3	325	26	US-60-339 453 358
6	221	10.3	1149	26	US-60-339 453 167
7	147	6.9	515	13	US-08-936-487-52

```

8 147 6.9 515 26 US-60-042-611-52
9 147 6.9 515 26 US-60-042-985-52
10 146 5.4 515 26 PCT-US01-18569-2412
11 79 3.7 164 21 US-09-758-449-1058
12 77 3.6 77 1 PCT-US01-00663-30151
13 77 3.6 77 22 US-09-864-761-37036
14 77 3.6 77 26 US-60-236-359-19127
15 72 3.4 135 21 US-09-757-028-2590
16 56 2.8 106 21 US-09-757-028-2541
17 53 2.5 176 21 US-09-757-028-1640
18 10 0.5 64 26 US-60-196-710-5862
19 10 0.5 64 26 US-60-196-712-3126
20 10 0.5 241 16 US-09-748-796-20694
21 10 0.5 241 26 US-60-096-409-29094
22 10 0.5 486 26 US-60-322-511-440
23 10 0.5 512 26 US-60-226-547-7
24 10 0.5 516 22 US-09-822-863-2
25 10 0.5 524 26 US-60-230-445-1772
26 10 0.5 524 26 US-60-230-445-1930
27 9 0.4 181 18 US-09-417-507-43564
28 9 0.4 448 24 US-10-015-127-13157
29 9 0.4 2096 20 US-09-614-150-23518
30 9 0.4 2147 26 US-60-167-217-22825
31 9 0.4 2147 26 US-60-173-464-18539
32 8 0.4 23 21 US-09-724-059-1234367
33 8 0.4 23 21 US-09-724-059-1234745
34 8 0.4 23 21 US-09-724-059-1235105
35 8 0.4 23 21 US-09-724-059-1235483
36 8 0.4 23 21 US-09-724-059-1235861
37 8 0.4 23 21 US-09-724-059-1236263
38 8 0.4 23 21 US-09-724-059-1236617
39 8 0.4 23 21 US-09-724-059-1236959
40 8 0.4 23 21 US-09-724-059-1237419
41 8 0.4 23 21 US-09-724-059-1237697
42 8 0.4 23 21 US-09-724-059-1242341
43 8 0.4 23 21 US-09-724-059-1242593
44 8 0.4 23 21 US-09-724-059-1242813
45 8 0.4 23 21 US-09-724-059-1243085

```

ALIGNMENTS

```

RESULT 1
US-09-603-665-5
: Sequence 5, Application US/09603665
: GENERAL INFORMATION:
: APPLICANT: Harty, Caroline
: APPLICANT: Rouqueleret, Lydie
: APPLICANT: Chumakov, Ilya
: APPLICANT: Cohen Alenine, Annick
: TITLE OF INVENTION: A NOVEL BAF28 GENE AND PROTEIN
: FILE REFERENCE: GENSET 063AUS
: CURRENT APPLICATION NUMBER: US/09/603,665
: CURRENT FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 60/141,323
: PRIOR FILING DATE: 1999-06-25
: PRIOR APPLICATION NUMBER: US 60/176,880
: PRIOR FILING DATE: 2000-01-18
: NUMBER OF SEQ ID NOS: 63
: SOFTWARE: Patent.pm
: SEQ ID NO 5
: LENGTH: 2144
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: 1694
: OTHER INFORMATION: Xaa-Ser or Asn
: NAME/KEY: VARIANT
: LOCATION: 1854
: OTHER INFORMATION: Xaa-Ala or Val
: NAME/KEY: VARIANT

```


XX WPI: 2001-63962/74.
DB N-PSD; AAS6647.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PI diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID No 32819; 101pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are a so used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (I) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAC00010-AAC30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at http://wipo.int/pub/published_pat_sequences.

SQ Sequence 68 AA:

Query Match 0.48; Score 8; DB 22; Length 68;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1655 KKKKGGEE 1662
|1111111
tb 29 KKKKGGEE 36

Search completed: July 2, 2002, 15:36:21
Job time: 384 sec

DE Drosophila melanogaster polypeptide SEQ ID NO 22518.
 XX Drosophila, developmental biology, cell signalling, insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-0504231.
 PF
 XX 23-MAR-2000; 2000RS-191637P
 PR
 XX 11-JUL-2000; 2000US-0614150.
 PP
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-050860/75.
 DR
 XX N-PSDB; ABL09345.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Disclosure, SEQ ID NO 22518, 21pp + Sequence Listing, English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176 ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 2096 AA;
 Query Match 0.4%, Score 9, DB 22, Length 2096,
 Best local Similarity 100.0%; Pred. No. 66;
 Matches 9; Conservative 0; Mismatches 0; Indels 0, Gaps 0.
 QY 30 LFDPKEAAT 38
 DB 31 lfdpkeaat 39
 |||||
 RESULT 14
 AAF70733
 ID AAF70733 standard; Protein; 20 AA.
 AC AAF70733;
 XX 26-APR-1991 (first entry)
 DT
 XX Sequence of leader polypeptide of human GM-CSF used in
 DE mammalian cell expression system for secretion of a polypeptide
 DE encoded of human granulocyte colony stimulating factor (hpc CSF)
 DE [Ala(1)].
 XX
 KW Haematopoietic disorders; therapy; aplastic anaemia;
 KW bone marrow transplant; burn wounds; leukaemia.
 XX
 OS Homo sapiens.
 XX
 FI Key Location/Qualifiers
 FT Peptide 1..17

FT Protein 1A 20
 XX WO8701132-A.
 PN
 XX 26-FEB-1987.
 PD
 XX 22-AUG-1986; 86WO-0001708.
 PF
 XX 03-MAR-1986; 86US-0835548.
 PR
 XX 23-AUG-1985; 85US-0768959.
 PP
 XX 23-AUG-1985; 85US-0768954.
 PA
 XX (KIRI-) KIRIN-AMGEN INC.
 PI (KIRI) KIRIN-AMGEN INC.
 XX Souza LM;
 PL
 XX WPI; 1987-064855/09.
 DR
 XX N-PSDB; AAN71092.
 DR
 XX Polypeptide with granulocyte colony stimulating factor activity
 PT - obtd. by recombinant DNA procedures for treating haematopoietic
 PT disorders
 PT
 XX Example, p52; 79pp; English.
 PS
 XX The examples describe procedures for the designing of probes for hpG-
 CC CSF cDNA and genomic clones, both of which are claimed. Specifically
 CC claimed are DNA sequences coding for (Ala 1)hpG-CSF; (Ser 36,42,64
 CC and 74)hpG-CSF and the corresp. Met-1 cpds. The novelty is that hpG-
 CC CSF is the prod. of procarcotic or eucarcotic expression of an
 CC exogenous DNA sequence. The construction of hpG-CSF vectors is also
 CC described in the examples.
 XX
 SQ Sequence 20 AA;
 Query Match 0.4%, Score 8, DB 8, Length 20;
 Best local Similarity 100.0%; Pred No 7 1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
 QY 1305 LLLLGTTVA 1312
 DB 6 llllgtva 13
 |||||
 RESULT 15
 ABG02460
 ID ABG02460 standard; Protein; 68 AA.
 AC ABG02460;
 XX 13-FEB-2002 (first entry)
 DT
 XX Novel human diagnostic protein #2451.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 DE food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2
 XX 11-OCT-2001
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PP
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI

QY 1420 FEQYVTKIVLAAAYGEK 1436
 ID AAM29882 standard; Protein: 77 AA.
 DB 61 feqyvktvlaaaygek 77

RESULT 11

AAM29882
 ID AAM29882 standard; Protein: 77 AA.

XX AC
 XX AC
 XX AC

XX 17 OCT 2001 (first entry)

XX Peptide #1919 encoded by probe for measuring placental gene expression.

XX Probe: microarray: human; placenta; antenatal diagnosis;
 genetic disorder

XX KW Homo sapiens.

XX OS

XX XX W0200157272 A2.

XX PN 09-AUG 2001.

XX PD 30-JAN-2001; 2001WO-US00661.

XX PF 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608409.

XX PR 03-AUG-2000; 2000US-0642366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human placenta.

XX PS Claim 27; SEQ ID No 3(151); 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENPs;
 see AAL1315-AA157545). The present sequence is a peptide encoded by one
 such probe. The probes are useful for producing a microarray for
 predicting, measuring and displaying gene expression in samples derived
 from human placenta. The probes are useful for antenatal diagnosis of
 human genetic disorders.

XX SQ Sequence 77 AA;

Query Match 4.6%, Score 77; DB 22; Length 77;
 Best Local Similarity 100.0%; Pred. No. 9.7e-68;

Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SUSGSIEVSRNVEIVVKIISVFVDALPHVPEHRRLPIVLVDTLGAKEFLWILLILL 1419
 DB 1 sdsqdsievstuvveevvkvkiiisvfdalphyvpehrripilvqlvdtlgaekflwlllll 60

QY 1420 FEQYVTKIVLAAAYGEK 1436
 DB 61 feqyvktvlaaaygek 77

RESULT 12

AAM05057
 ID AAM05057 standard; Protein: 77 AA.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX 09-OCT-2001 (first entry)
 XX Peptide #3739 encoded by probe for measuring breast gene expression.
 XX Probe: human; breast disease; breast cancer; development disorder;
 inflammatory disease; proliferative breast disease; non carcinoma tumour.
 XX KW Homo sapiens.
 XX OS
 XX W0200157272 A2.
 XX PN 04-AUG-2001.
 XX PD 29-JAN-2001; 2001WO-US00661.
 XX PF 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608409.
 XX PR 03-AUG-2000; 2000US-0642366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PR WPI; 2001-476286/51.

XX PT Novel single exon nucleic acid probe used to measuring gene expression
 in a human breast.

XX PS Claim 27; SEQ ID No 13797; 323pp; English.

XX CC The present invention relates to novel single exon nucleic acid probes
 (see AAL0010-AA10067). The present sequence is a peptide encoded by one
 such probe. The probes are useful for measuring human gene expression in
 a human breast sample, where the probe hybridises at high stringency to a
 nucleic acid expressed in the human breast. The probes are useful for
 predicting, diagnosing, grading, staging, monitoring and prognosis
 diseases of the human breast, particularly those diseases with polygenic
 aetiology. The diseases include: breast cancer, disorders of development,
 inflammatory diseases of the breast, fibrocystic changes, proliferative
 breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WPI
 at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 77 AA;

Query Match 3.6%, Score 77; DB 22; Length 77;
 Best Local Similarity 100.0%; Pred. No. 9.7e-68;

Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SUSGSIEVSRNVEIVVKIISVFVDALPHVPEHRRLPIVLVDTLGAKEFLWILLILL 1419
 DB 1 sdsqdsievstuvveevvkvkiiisvfdalphyvpehrripilvqlvdtlgaekflwlllll 60

QY 1420 FEQYVTKIVLAAAYGEK 1436
 DB 61 feqyvktvlaaaygek 77

RESULT 13
 ID ABB65242
 AC ABB65242

XX AC ABB65242;
 XX AC ABB65242;

XX DT 26-MAR-2002 (first entry)

XX XX

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.

XX
SQ Sequence 77 AA;

Query Match 3.6%; Score 77; DB 22; Length 77;
Best Local Similarity 100.0%; Pred. No. 9.7e-68;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SDSGDSIEVSRNVEEIVVKIISVFVDALPHVPEHRRPLILVOLVDTLGAEKFLWILL 1419
DB 1 sdsdgsievsrnveeiivvkiisvfvdaiphvpehrrplilvolvdtlgackflwlllll 60

QY 1420 FEQYVTKTVLAAAYGEK 1436
DB 61 feqyvtktvlaaaygek 77

RESULT 9
AAM69542
ID AAM69542 standard; Protein: 77 AA.
XX
AC AAM69542;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 29848.

XX Human; bone marrow expressed exon, gene expression analysis, probe,
KW microarray; cancer, leukemia, lymphoma, myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180412.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR,
XX
XX WPI: 2001 488900/53.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4: SEQ ID NO: 29848; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.

XX Sequence 77 AA;

Query Match 3.6%; Score 77; DB 22; Length 77;
Best Local Similarity 100.0%; Pred. No. 9.7e-68;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SDSGDSIEVSRNVEEIVVKIISVFVDALPHVPEHRRPLILVOLVDTLGAEKFLWILL 1419
DB 1 sdsdgsievsrnveeiivvkiisvfvdaiphvpehrrplilvolvdtlgackflwlllll 60

QY 1420 FEQYVTKTVLAAAYGEK 1436
DB 61 feqyvtktvlaaaygek 77

RESULT 10
AAM17368
ID AAM17368 standard; Protein: 77 AA.
XX
AC AAM17368;
XX
DT 12-OCT-2001 (first entry)

XX Peptide #3802 encoded by probe for measuring cervical gene expression.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX Homo sapiens.
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-488901/53

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 27: SEQ ID NO 22194; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
XX (SENPs; see AAT10068-AA128459) The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences

XX Sequence 77 AA;

Query Match 3.6%; Score 77; DB 22; Length 77;
Best Local Similarity 100.0%; Pred. No. 9.7e-68;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SDSGDSIEVSRNVEEIVVKIISVFVDALPHVPEHRRPLILVOLVDTLGAEKFLWILL 1419
DB 1 sdsdgsievsrnveeiivvkiisvfvdaiphvpehrrplilvolvdtlgackflwlllll 60

```

PT Human genome derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
XX
PS Claim 27: SEQ ID No 29019; 53pp - sequence listing; English
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human fetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 77 AA:

Query Match 3.6%; Score 77; DB 22; Length 77;
Best Local Similarity 100.0%; Pred. No. 9.7e-98;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SDSGDSIEVSNVERIVVKIIISVFVDALPHVPEHRLPILVQLVDTLGAKEFLWIIILL 1419
DB 1 sdsqdsievsnvceivvkiisvfdalphehrfripilvqlvdtlgaekflwiiill 60

QY 1420 FEQYVTKIVLAAAYGEK 1436
DB 61 feqyvktivlaaaygek 77

RESULT 7
AAB21748
ID AAB21738 standard; Protein: 77 AA.
AC AAB21748;
XX
DT 23-JAN 2002 (first entry)
XX
DE Protein #4747 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
FN W0200157274 A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US02466.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WP: 2001-488894/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
PS Claim 15; SEQ ID No 23508; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for

```

```

CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535 ABA1305). The present sequence is a protein encoded by one such
CC gene. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 77 AA:

Query Match 3.6%; Score 77; DB 22; Length 77;
Best Local Similarity 100.0%; Pred. No. 9.7e-98;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1360 SDSGDSIEVSNVERIVVKIIISVFVDALPHVPEHRLPILVQLVDTLGAKEFLWIIILL 1419
DB 1 sdsqdsievsnvceivvkiisvfdalphehrfripilvqlvdtlgaekflwiiill 60

QY 1420 FEQYVTKIVLAAAYGEK 1436
DB 61 feqyvktivlaaaygek 77

RESULT 8
AAM57146
ID AAM57146 standard; Protein: 77 AA.
AC AAM57146;
XX
DT 05 NOV 2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID No: 29251.
XX
KW Human; brain expressed exon; gene expression analysis; proteo-
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
FN W0200157275 A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WP: 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
PS Example 4; SEQ ID No: 29251; 650pp - Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system

```

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primers sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAR02446 to
 CC AAR95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 349 AA;

Query Match 5.9%; Score 127; PR 22; Length 349;
 Best Local Similarity 100.0%; Pred. No. 1, 1, 116;
 Matches 127; Conservative 0, Mismatches 0, Indels 0, Gaps 0;
 QY 2018 ALMPLVDQLENLGGSEKFEKVTIKLILICIAQFSVAMADUSLWKPILNYOILLKTRDSS 2077
 DB 223 almpdvqlecnrlygceklqetvklilpelclafsfvamadslwkpilnyqlliktrdss 282
 QY 2078 PKVFEALITVLALAEKKNYIVLPESIFFLAEIEMEDELVEVEHQVQKTIQQLETVLG 2137
 DB 283 pk-rfaallitvlalaeakknynivlpesifflaelmedeceevehqckliqqlletvlq 342
 QY 2138 EPLQSYF 2144
 DB 343 eplqsyf 349

RESULT 5
 ABB31185 ID ABB31185 standard; Peptide; 77 AA.
 XX
 AC ABB31185;
 DT 01-FEB-2002 (first entry)
 XX
 DE Peptide #386 encoded by breast cell single exon nucleic acid probe.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-0500662
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0642006.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236459.
 PR 04-OCT-2000; 2000US-0024266.
 XX

XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-496933/54.
 XX

PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -

XX
 PS Claim 27; SEQ ID NO 14153; 327bp; sequence listing, English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BI 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 77 AA;

Query Match 3.6%; Score 77; PR 22; Length 77;
 Best Local Similarity 100.0%; Pred. No. 9, 7e-68;
 Matches 77; Conservative 0, Mismatches 0, Indels 0, Gaps 0;
 QY 1360 SDSGSS:EVSRNVEEIVVKLIISVFVDALHYVPHRRRIPIILVQIDTCARKEFWILLILL 1419
 DB 1 sdsqdsievsrnrveeivvklisvfvdalhyvphrrripililvqidiqackfwiwillll 60
 QY 1420 FEQYVTKTVLAAAYGEK 1436
 DB 61 feqyvtktvlaaaygek 77

RESULT 6
 ABB36384 ID ABB36384 standard; Peptide; 77 AA.
 XX
 AC ABB36384,
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #3890 encoded by human foetal liver single exon probe.
 XX
 KW Human, foetal liver, gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-0500669
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0642006.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000US-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001 483447/52.
 XX

QY 1468 QYLLKPKPEKEETIPKAVSPKSKSQRMLQVFNWFTHTSKQLHPKPLSVSPMSCLISS 1527
 |||||
 Db 1382 QYLLKPKPEKEETIPKAVSPKSKSQRMLQVFNWFTHTSKQLHPKPLSVSPMSCLISS 1441
 QY 1528 NPLKVVVFSPFELLKGLFPFPIETVLYTSANALSWEPNAPFIVVFWFALSPATL 1587
 |||||
 Db 1442 NPLKVVVFSPFELLKGLFPFPIETVLYTSANALSWEPNAPFIVVFWFALSPATL 1501
 QY 1588 LKYNALLPTETFPVIRGLVGNELPSVRRKALDLNNKLNQNSWKKTIIVTRFKLVPD 1647
 |||||
 Db 1502 LKYNALLPTETFPVIRGLVGNELPSVRRKALDLNNKLNQNSWKKTIIVTRFKLVPD 1561
 QY 1648 LLAIVQK 1655
 |||||
 Db 1562 LLAIVQK 1569

RESULT 3
 AAW54099 standard; Protein: 515 AA.

AAW54099:

28 SEP-1998 (first entry)

Homo sapiens HAP28 sequence.

BAPD1: ring protein; BRCA1; breast cancer; risk; diagnosis.

Homo sapiens.

W09812427-A2.

26 MAR-1998.

19 SEP-1997; 97W02516842.

04 APR-1997; 97US-0142985.

20 SEP-1996; 96US-0125296.

03 APR-1997; 97US-0142611.

(TEXA) UNIV TEXAS SYSTEM.

Hart R, Howcock AM.

WPI: 1998 240417/20

N-PSR: AAV24135

DNA sequence encoding HAPD1, B123, BE2, BE14, BE31 or BE445 - which

as breast cancer antigen, HRCAL, binding proteins are useful to

identify patient having or at risk of developing cancer

Disclosure: Page 287-288; 348pp; English.

The sequence is that of a protein which can be used in the preparation of the recombinant breast cancer antigen, HRCAL, binding proteins BAPD1, B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a BAPD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence specifically a wild type BAPD1 composition for the detection or purification of HRCAL, useful to identify a patient having, or at risk of developing cancer. HAPD1 can be used in the preparation of an anti-HAPD1 antibody, and in the detection and purification of a HRCAL protein. BAPD1, B123, BE2, BE14, BE31 or BE445 can be used in the identification of a binding protein agonist or antagonist that alters the binding of BAPD1, B123, BE2, BE14, BE31 or BE445 to BRCAL or the biological activity of the HRCAL-HAPD1, B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to detect HAPD1, B123, BE2, BE14, BE31 or BE445, a specific anti-HAPD1 antibody can be used to identify a patient having or at risk of developing cancer

Sequence 515 AA.

Query Match 8.0%; Score 171; DB 19; Length 515;
 Best Local Similarity 99.6%; Pred. No. 3,66-160;
 Matches 271; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1655 TAVELIATPEPEERNNVLSALUCIAEVISTLEALALPQLFSLMFSLITIMPINSELVSSE 1754
 |||||
 Db 66 TAVELIATPEPEERNNVLSALUCIAEVISTLEALALPQLFSLMFSLITIMPINSELVSSE 125
 QY 1755 VYLLSALAALQKVVETLPHFISPYLEGILSOVHLEKIIISEMSASQANIRLTSLKKTILA 1814
 |||||
 Db 126 VYLLSALAALQKVVETLPHFISPYLEGILSOVHLEKIIISEMSASQANIRLTSLKKTILA 185
 QY 1815 VYLLAPRVLLPAIKYVYKQIPKWKNNHMGPFMSIIQPHIGXMKKKEETISQSOQTAPFIPA 1874
 |||||
 Db 186 VYLLAPRVLLPAIKYVYKQIPKWKNNHMGPFMSIIQPHIGXMKKKEETISQSOQTAPFIPA 245
 QY 1875 LDFRACHSFNOLFEVKTENCIIIDVIVAVVVLSEVTPPEPLFFKLTWAKIEDAPKDRLL 1944
 |||||
 Db 246 LDFRACHSFNOLFEVKTENCIIIDVIVAVVVLSEVTPPEPLFFKLTWAKIEDAPKDRLL 305
 QY 1945 IFYNLADGIAEKIKGLFTLFGHLYKPPFADTL 1966
 |||||
 Db 306 IFYNLADGIAEKIKGLFTLFGHLYKPPFADTL 337

RESULT 4

AAW92729

ID AAW92729 standard; Protein: 345 AA.

AAW92729:

26 JUN-2001 (first entry)

Human protein sequence SEQ ID NO:11159.

Human, primer, detection, diagnosis, antisense therapy; gene therapy.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000JP 0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0302531.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0184767.

09-JUN-2000; 2000JP 0241859.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sutoyama T, Wakamatsu A, Naito K, Otsuki I;

WPI: 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 11159; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

Db 1681 fgaenqefvfvxtavktlaperkeeknvlgsatctacavtstleatlapqpslmsl 1740
 QY 1741 LTTMKNTSELVSSEVLLSALALOKWWTLPHPFISPYLEGILSYVILHLEKLTSEMSSAS 1800
 Db 1741 LTTMKNTSELVSSEVLLSALALOKWWTLPHPFISPYLEGILSYVILHLEKLTSEMSSAS 1800
 QY 1801 QANTRITSLTKKTLATTLAPRVLIPALIKKTYKOJEKNMKNHMGPPMSILQSHIGXMKKEEL 1860
 Db 1801 qantritslktlatlprvllpalktkyqekwknhmgppmsilqshixmkkeel 1860
 QY 1861 TSHQSULTAFLEALDFRAHSENDLEEVGKTENCIICDLVAMVVKUSEVTFPLPKLF 1920
 Db 1861 tshqsultaflealdfraghscndleevgktencicdlvavmkvisevtfplpklf 1920
 QY 1921 DWAKTADPKDELLTFYNLADLAELKLGLETLFAGHLVKEPALTLXVNI SKTGPAPFP 1980
 Db 1921 dwaktadpkdrlltfynladlaelklglgtlflaghlvkepalatlkvni sktgskdeafid 1980
 QY 1981 SENHPRKCCCLLQFLNCLYKIFLFDTHFISKPRAXALMMPIDVDGLENRIJGGEKFOER 2040
 Db 1981 sendpekcclllqflncllykiifldthfiskpraxalmpidvdqlenrlggeeikfer 2040
 QY 2041 VTKHLHCAQFSVAMADDSLWKLNVQILLTKDSSPKVPFPAALITVLALAPKLKENYI 2100
 Db 2041 vtkhlpclqfsvamadslwklpnyqilltktdsspkvrfpaalilcvlalaacklkenyi 2100
 QY 2101 VLLPESIPFLAELMEDCEVEHQCKTQIQLETLVLGEPLQSYF 2144
 Db 2101 vllpesipflaelmedcevehqcktiqqletvlgeplqsyf 2144

RESULT 2

ABG15232

ID ABG15232 standard; Protein; 1569 AA.

XX AC ABG15232;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #15232

XX KW Human, chromosome mapping, gene mapping, gene therapy, forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN W0200175067-A2.

XX PU 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-0808631.

XX PP 31-MAR-2000; 2000MS-0540217

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSB-) HYSBQ INC.

XX PI Dramatic RT, Liu C, Tang YT.

XX DR WPI, 2001 639362/73.

XX DR N-PSDB; AAS79419.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity -

XX PS Claim 20; SEQ ID NO 45591; 103pp, English.

XX CC The invention relates to isolated polynucleotide (I) and

XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX CC and gene mapping, and in recombinant production of (I). The

CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_prot_sequences.

XX SQ Sequence 1569 AA.

Query Match 42.4%, Score 908; Db 12; Length 1569;

Best Local Similarity 100.0%; Pred No. 0;

Matches 908; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 EIPSEWHIELMDRGPVEIMAHYVVEHINSTRVAVENSVPLFSLKKFIYALAKPSHP 807
 Db 662 eipsewhielmdrqpveimahyveehinstqvavedsvlflskkfiyalkapskip 721
 QY 808 KGDIMWNPQGLKEDSDRYLHLILGLFEMMLNCAQAVHPRVLMKILFKVHLKDVLPKPKC 867
 Db 722 kgdimwnpeqikedrlyhlilglfemmlngadavhprvlmkifkvhldvlpkpkc 781
 QY 868 SVIWTYGSLSNPLNCSVKTVLTQALYVYVAMLSQSKTYKKHGLASTSSPVSLLNL 927
 Db 782 svlwtgsslsnplncsvktvltqalyvyvamlssqsktykkhglastsspvsllnl 841
 QY 928 GSPVKEVRRRAIQCLOALSGVASPPYLIIDHLISKAERITSDAAVYIQDLATLPEELOPE 987
 Db 842 gspvkevrriaalqclqalsgvaspfiylihhliskaeilttdaayviqdlatlpeelpe 901
 QY 988 KKLKSHOKLSLTKNLLSCVYSCPSYIAKOLMKVIOGVNGEMVLSQILPMAPQLIKETOK 1047
 Db 902 kklkshoklseltknllscvyscpsyiaaklmkviogvngemvlsqilpmaeqilekikq 961
 QY 1048 EPTAVLKDAMVLIHLITGKYNEFSVSLNEDPKSLDIFIKAVHTTKELYAGMPTQITATL 1107
 Db 962 eptavlkdeamvlihtlgkynefsvslinedpksldifikavhttkelyagmptqital 1021
 QY 1108 EKLTTPPFAATSDBKVQOKLLRMILFDIALVNCKNSHCAOTVSSVPKGISVNAPOVRIRLEP 1167
 Db 1022 eklttppfaatfdbkvokllrmilfdialvncknsahcaotvssvpkqisvnaeqvrielep 1081
 QY 1168 PKAKAPDGTIVQKKPKKMKQKKKSQDFPSVQVGVGVSYWDPVTHLPLELQHKKKIRSPQILV 1227
 Db 1082 pdkakpigtvqgktrqkmqgkksqdfpsvqvsgvsgywdvthllplelqhkktirspqilv 1141
 QY 1228 FTLENLLSEGLEFPYQELQNMETKQLTILSLNLNLTGKLSKSLGSKIPKDIIDEEFENVEL 1287
 Db 1142 ftlenllseglefpyqelnmetkqltillsnlntgklskslsgskipkdildeeefenvel 1201
 QY 1288 IVQCIRLSEMPQTHHHLILDLGTAVGTFPKDKVLIHNMISITPMGANVMHFDIPTYSPQVIN 1347
 Db 1202 ivqcirlsempqthhhlildlgtavgtfpkdkvlihnmsitpmganvmhfdiptyspqvin 1261
 QY 1348 KTVKMWIPALIQSDSGDSIFVSRNVEEIVVKILISVVFVIALPHVPEHRLPILVQLVDILG 1407
 Db 1262 ktvkmwipaligsdsgdsifvsrnveeivvkiilissvfvialphvpehrlpilvqlvdlilg 1321
 QY 1408 AEKFWILLIILFEQYVTKTVLAAAYGEKDAILEADTEFWFSVCCESVQHQIQSLMNL 1467
 Db 1322 aeckfwilliilfedqyvtktvllaaygekdaileadtefwfsvccesvqhqiqlsmnl 1381

XX Harry C. Bouquelerot L. Chumakov L. Cohen, Akenine A.
 XX WPI: 2001-367042/38.
 DR N PSDB: AAF84909, AAF84910.
 XX New BAP28 polypeptides and polypeptides overexpressed in prostate
 PI cancer cells for diagnosing prostate tumors, e.g. by hybridization or
 PT polymerase chain reaction assays.
 XX Claim 14; Page 297-304; 349pp; English.
 PS
 XX the invention is directed to BAP28 polypeptides, BAP28 polynucleotide
 CC sequences and regulatory regions located at the 3' and 5' ends of the
 CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard
 CC recombinant methodology. BAP28 polynucleotides and polypeptides have been
 CC found to be over expressed in prostate tumor cells, therefore levels of
 CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
 CC reaction (PCR)) to diagnose patient suffering from or susceptible to
 CC prostate cancer. Antibodies specific for the BAP28 polypeptides are
 CC useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
 CC useful in genetic analysis. The present sequence represents a protein
 CC encoded by a first cDNA sequence of the BAP28 gene consisting of the
 CC exons 1 to 45.
 XX
 XX Sequence 2144 AA:
 SQ
 Query Match 99.8%; Score 2140; DB 22; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTSIAQIQRLALPQSDASLSRFEVASLFDPEKATINPTAFATGCTGLEELLIDP 60
 Db 1 mtsiaqiqrlalpqsdaslsrfevaslfdpekatinptafatgctgleellidp 60
 QY 61 SFEQFEAPLPSQAKTLERSVOTKAVKQIDENLSLPLHLSPFLKPAQCKLFWLHR 120
 Db 61 sfeqfeaplpssqaktlersvotkavkqidensllplhlspflkpaqcklewlhr 120
 QY 121 FHILYQDLSLAVLPEYHETFPVRVLTLKINNSKHPFWLLPVLKSGVPLAKGLIT 180
 Db 121 fhilyqdslslavlpeyhettvrvrltlkinnshkhrfwllpvksgvplakglit 180
 QY 181 HCYKDLGEMDFICSLVTKSVKFAEYPESSAQLVLLAFAYASTIVSALVAAPERVSNTIA 240
 Db 181 hcykdldgmdficslvtksvkfayepessaqllvllafayastivsalvaapervsntia 240
 QY 241 KLPFYIQGLKSSLPDYKAAIYMLCOISVKVIMENIFVNSIASQITKLIKPSLKDG 300
 Db 241 klpfyiqglksslpdykkaaitymlcoisvkvimnifvnsiasqitklikpslkdg 300
 QY 301 LSCILIVLLORQKPSLCKKPPHLCNVPDLITILGISETYDVSPLLYMLPHLWSTIH 360
 Db 301 lscilivllorokpslckkpphlcnvpldlitilgiseytydvspllymlphlwstih 360
 QY 361 HVTGEETGEMGQIKYKRELAITKLSLKNLDHLLSLPFEYISYSSQEMDSNKVSL 420
 Db 361 hvteetgmgqikykreleaitklslnldhllslpfeeyisyssqemdsnkvs 420
 QY 421 INQFQPLIRLESKYPRTHLVLEHHEKTAADLKQELPHQFVSLSLTSQGYQFLAUSD 480
 Db 421 inqfqpilirleskyprthlvleehhektadlkqelphqfvs sltsqgyqflausd 480
 QY 481 TSLMLSLNIHLAPVILAMNHLKIMTSGVDSPEFKEAVLAPLGDNDTDDVLSAISA 540
 Db 481 tslmlslnihlapvilamnhlkimtsgvdspefkeavlaplgdndtdvlsaisa 540
 QY 541 FEIFKEHFSSEFTISNINLPQAFELSKNFEWYFLKTAADITKEETISENDLSNOVV 600
 Db 541 feifkehfsseftisnlnlpqafelsknfeywylktaaditkeetisendlsnovv 600
 QY 601 VCIIPVWVNNDDTESAPMKIATIVISQSGICSLHPLIPQWFALENVTKSPKELFGVA 660
 Db 601 vciipvwvnnddtesapmkiativisqsgicslhplipqwfalenvtpkelpkelfgva 660

Db 601 vciipvwvnnddtesapmkiativisqsgicslhplipqwfalenvtpkelpkelfgva 660
 QY 661 NOKMTELLADNINLGDPSMILKVEDLISVGEESFNLIKQKVTPIVILSVLVSQVSSLIKE 720
 Db 661 nokmteelladninlgdpsmilkvedlissvgeesfnlikqkvtpivilsvlvsqvs 720
 QY 721 THPFAIIVPVSLLQKKIKKLESVITAVEIPSHWHIDLMEDRCIFVIEWAHYVHEHINJOR 780
 Db 721 thpfaiivpvsllqkkikklesvitaveipshwhidlmmedrcifvievahyveehin 780
 QY 781 VAVEDSVPLVPSLAKFIYALAKAPSFPKCDIWNPNQIKLSDRYLHLLIGLIFPMMLNCA 840
 Db 781 vavedsvplvpslakkfiyalakapsfpkcdiwnpnqiklsdryllhlliglifpmmln 840
 QY 841 IAVHFRVLMKLFIRVHLEDFVLFKFCGSLVTGYSSLSNPLNCVKTVLOUALYVSGCAM 900
 Db 841 iavhfrvlmklfirvhledfvlfkfcgslvtgysslsnplncvktvloualyvs 900
 QY 901 LSSUKTQCKHQULASISSVWVLSLILNAGSPKVKVRAALQCLQALSGVASRYYLIDHLL 960
 Db 901 lssuktqckhqulasssvwvlsllilnagspkvkvrallqclqalsgvaspyliidh 960
 QY 961 SKAPEITSDAAVYIQLDITLPEELQRPKKIKSHQKLSFTLKNLSCVYSCPSYIAKDIMK 1020
 Db 961 skapeitdsaaavyiqlditlpeelqrpkkikshqklsftlknllscvyscpsy 1020
 QY 1021 VLGIVNEMVLSQLPMAEQULEKIQEPTAVLKUDEAMVLHLITLGYNEFSVSLNEDPK 1080
 Db 1021 vlgivngmvlsqllpmaequllekiqueptavlkudeamvhlitlgynefsvsl 1080
 QY 1081 SLDIFIKAVHITKELYACMPTIQTALFKITKPPFAALISDEKVOOKLILHMLFDLLVNCN 1140
 Db 1081 sldifikavhitkelycmptiqtalfkittppfaalisdokvooklilhmlfdllvnc 1140
 QY 1141 SHCAQTVSSVFKGIVSNAPQVPTLPFPQKAKPICTVQVQPPQVCKQKSSQINFSWQVVC 1200
 Db 1141 shcaqtvsstvfkgisvnapqvptlpfpqkakiptvqvqpqvckqkssqinfs 1200
 QY 1201 TSYQWQVTLILELQHKKKLPS-PULLVPTLENLS-SPULPELPUEGSKNEYTKQLILSLL 1260
 Db 1201 tsyqvwvtillelqhkkkllps-pullvptlenls-spulpelepuegskneytk 1260
 QY 1261 NFOCKISPDGCKTDPKDIIDEEKFNVELIVQCIRISEMIQTHHALLILGTAVIPDKVL 1320
 Db 1261 nfockispdgcktdpkdiideekfnvelivqcirisemiqthhallilgtavipdk 1320
 QY 1321 HNIMSFTFMGCVNVRDLDTYSFQVINKTKMVIPALIQSDSGDSIFVSRNVEFTVVK 1380
 Db 1321 hnimsftfmgcgvnvrlddty sfqvinktkmvipalioqsdsgdsifvsrnveft 1380
 QY 1381 SVEVDALPHVPEHRLPLTVOLVDTLGAKEFLWILLILLFEOYVTKTLAAAYGEKDAIL 1440
 Db 1381 svevdalphvpehrlpltlvolvdtlgaekflwllillllfeyovtktlaaayge 1440
 QY 1441 PADYKFWPSVOCRRFSVQHQIQELMNIIOYLKLPKEKETIPKAVSPNKSFSQEMIGVF 1500
 Db 1441 eadylewpsvccelssvqhqislmiloyliklpeekeelpkavstnksesqemil 1500
 QY 1501 NVETHSKQLRHPKPISVSPMQLISSNNFIPKVVFSQGPPIKGLPPELLETIVIGYISA 1560
 Db 1501 nvethskqlrhpkipisvpsmqllssnnfipkvvfsqgppikglppeleltivigy 1560
 QY 1561 VAQSMERNADKLTVKFWPALLSKAYDILDKVNALLPTETFPVIVKVLVNPPLPSVPPKAL 1620
 Db 1561 vaqsmernadkltvkfwpallskaydildkvnallptetfpvivkvlvnpplpsvp 1620
 QY 1621 DILNNKADQUNISWKKIIVIRFKIYVPLDIAIVORAKKSGEERQAINQOITATYTLKCKN 1680
 Db 1621 dilnnkadquniswkkiiwirfkiyvpldiaivorakksgeerqainqoitatytl 1680
 QY 1681 PGAEHNDPDPVIVLTVAVKLIAIFERKEKNVIGSALICIAEVTITLPAIADPOLSLMPSL 1740
 Db 1681 pgae hndpdpvivl tvavkliaiferkeknvigsaliciaevtitlpaiadpolslm 1740

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 2, 2002, 15:29:57 ; Search time 52.51 seconds
(without alignments)
4545.178 Million cell updates/sec

Title: US-09-603-665-5
Sequence: 1 MSLAAGQIGRIALHDSVASL

CORRIGIQUVIGPPIQSYF 2144

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 11107396 residues

Word size : 6

Total number of hits satisfying chosen parameters: 11424

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_028202.*

- 1: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/qqqdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2140	99	8	2144	22 AAR6029
2	608	42	4	1569	22 AAR6029
3	171	8	0	515	19 AAR54099
4	127	5	9	349	22 AAR92729
5	77	3	6	77	22 AAR31185
6	77	3	6	77	22 AAR36384
7	77	3	6	77	22 AAR21738
8	77	3	6	77	22 AAR57146
9	77	3	6	77	22 AAR69542
10	77	3	6	77	22 AAR17368
11	77	3	6	77	22 AAR20882

12	77	3	6	77	22	AAR05057
13	2096	22	22	22	22	AAR65242
14	8	4	4	20	8	AAP70743
15	8	4	4	68	22	ABG02460
16	8	4	4	78	22	AAH55838
17	8	4	4	78	22	AAH57629
18	8	4	4	78	22	AAH58034
19	8	4	4	93	22	AAH10562
20	8	4	4	93	22	AAH92998
21	8	4	4	107	22	ABE65755
22	8	4	4	108	22	ABE96435
23	8	4	4	142	10	AAH94855
24	8	4	4	142	17	AAW03933
25	8	4	4	144	7	AAH60545
26	8	4	4	144	7	AAH60418
27	8	4	4	144	8	AAH70656
28	8	4	4	144	8	AAH70657
29	8	4	4	144	9	AAH80527
30	8	4	4	144	9	AAH81886
31	8	4	4	144	9	AAH81885
32	8	4	4	144	11	AAH04098
33	8	4	4	144	16	AAH71118
34	8	4	4	144	17	AAW03932
35	8	4	4	144	17	AAW03934
36	8	4	4	144	17	AAH92800
37	8	4	4	144	19	AAH60041
38	8	4	4	144	21	AAH18633
39	8	4	4	144	21	AAH20002
40	8	4	4	144	22	AAH84602
41	8	4	4	144	22	AAH50871
42	8	4	4	174	22	AAH92186
43	8	4	4	174	22	AAH64138
44	8	4	4	179	22	AAH64492
45	8	4	4	179	22	AAH85482

ALIGNMENTS

RESULT 1	
AAR85029	
10 AAR85029 standard, protein; 2144 AA.	
XX	
XX AAR85029;	
XX	
XX	on-AAR85029 (first entry)
XX	Protein encoded by AAR28 cDNA consisting of exons 1 to 45.
XX	AAR28, prostate, tumour; cancer; diagnostic, genetic analysis.
XX	Homo sapiens.
XX	
XX	Key Location/Qualifiers
XX	Misc-difference 1694
XX	Misc-difference /label= Ser or Asn
XX	Misc-difference 1854
XX	Misc-difference /label= Ala or Val
XX	Misc-difference 1967
XX	Misc-difference /label= Asp or Asn
XX	Misc-difference 2017
XX	Misc-difference /label= Gly or Glu
XX	
XX	W0200100064-A2
XX	
XX	04-JAN-2001.
XX	
XX	23-JUN 2000; 2000WO-1801183.
XX	
XX	25-JUN-1999; 9905-0141323.
XX	18-JAN-2000; 2000US-0176880.
XX	
XX	(GSI) GENSET.

Peptide #3739 enco
Drosophila melanog
Sequence of leader
Novel human diapo
PDZ encoded domain
Human tax interact
human tax interact
Human pancreatic c
Human digestive sy
Drosophila melanog
Furcillo P. abyssal
Human granulocyte
GM-CSF (N- and O-1
Colony stimulating
Human granulocyte
Sequence of a huma
Granulocyte macrop
Sequence of human
Metaprotease
GM-CSF encoded by
GM-CSF (N-linked s
GM-CSF (one N-link
Human GM-CSF. Hom
Human granulocyte
Amino acid sequen
Human granulocyte-
Amino acid sequen
Human GM-CSF. Hom
Human protein sequ
Human secreted pro
Human secreted pro
Human APP-like pro



Wed Jul 3 12:05:23 2002

Db 10 HMKPFA 16

```

RESULT 15
US-10-041-018-309
; Sequence 309, Application US/10041018
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Bart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent version 3.1
; SEQ ID NO 109
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-309

```

```

Query Match      3.5%; Score 7; DB 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 41 HMKPFA 47
 Db 10 HMKPFA 16

Search completed: July 2, 2002, 16:16:47
 Job time: 824 sec

Sequence 8006, Application US/60-60039
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xianrong
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
: FILE REFERENCE: 48-1052052A
: CURRENT APPLICATION NUMBER: US/60-60039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 8006
: LENGTH: 852
: TYPE: PRT
: ORGANISM: Rhodobacter sphaeroides
US-60-60039-8006

Query Match 3.5%: Score 7; DH 7; Length 852;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VLALAEK 178
DB 474 VLALAEK 480

RESULT 11
US-60-60039-9821
: Sequence 9821, Application US/60-60039
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xianrong
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS WITH IMPROVED PROPERTIES
: FILE REFERENCE: 48-1052052A
: CURRENT APPLICATION NUMBER: US/60-60039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 9821
: LENGTH: 894
: TYPE: PRT
: ORGANISM: Desulfitobacterium halniense
US-60-60039-9821

Query Match 3.5%: Score 7; DH 7; Length 894;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VLALAEK 178
DB 508 VLALAEK 514

RESULT 12
US-10-041-018-199
: Sequence 199, Application US/10041018
: GENERAL INFORMATION:
: APPLICANT: Matsuda, Seiichi P.T.
: APPLICANT: Hart, Elizabeth A.
: TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
: FILE REFERENCE: P020800S/10025547
: CURRENT APPLICATION NUMBER: US/10-041-018
: PRIOR FILING DATE: 2002-01-07
: PRIOR APPLICATION NUMBER: US 60/259880
: NUMBER OF SEQ ID NOS: 41

: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 199
: LENGTH: 1045
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-041-018-199

Query Match 3.5%: Score 7; DH 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 HLVKPFA 47
DB 10 HLVKPFA 16

RESULT 13
US-10-041-018-218
: Sequence 218, Application US/10041018
: GENERAL INFORMATION:
: APPLICANT: Matsuda, Seiichi P.T.
: APPLICANT: Hart, Elizabeth A.
: TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
: FILE REFERENCE: P020800S/10025547
: CURRENT APPLICATION NUMBER: US/10-041-018
: PRIOR FILING DATE: 2002-01-07
: PRIOR APPLICATION NUMBER: US 60/259880
: PRIOR FILING DATE: 2001-01-05
: NUMBER OF SEQ ID NOS: 413
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 218
: LENGTH: 1045
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-041-018-218

Query Match 3.5%: Score 7; DH 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 HLVKPFA 47
DB 10 HLVKPFA 16

RESULT 14
US-10-041-018-264
: Sequence 264, Application US/10041018
: GENERAL INFORMATION:
: APPLICANT: Matsuda, Seiichi P.T.
: APPLICANT: Hart, Elizabeth A.
: TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
: FILE REFERENCE: P020800S/10025547
: CURRENT APPLICATION NUMBER: US/10-041-018
: PRIOR FILING DATE: 2002-01-07
: PRIOR APPLICATION NUMBER: US 60/259880
: PRIOR FILING DATE: 2001-01-05
: NUMBER OF SEQ ID NOS: 413
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 264
: LENGTH: 1045
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-041-018-264

Query Match 3.5%: Score 7; DH 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 HLVKPFA 47

09-603-665-5_copy_1917_2117_olig6.rapn

Wed Jul 3 12:05:23 2002

```

Query Match      3.5%  Score 7; DB 7; Length 393;
Best Local Similarity 100.0%; Pred. No. 59;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 LAEKLE 181
    |||||
Db 272 LAEKLE 278

RESULT 8
US-10-155-881-27684
; Sequence 27684, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Iutifiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 3/595
; SEQ ID NO 27684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-27684

Query Match      3.5%  Score 7; DB 6; Length 409;
Best Local Similarity 100.0%; Pred. No. 61;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 TVLALAE 177
    |||||
Db 335 TVLALAE 341

```

```

RESULT 9
US-60-360-039-2667
; Sequence 2667, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 4/374
; SEQ ID NO 2667
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Synecocystis sp.
US-60-360-039-2667

```

```

Query Match      3.5%  Score 7; DB 7; Length 415;
Best Local Similarity 100.0%; Pred. No. 62;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 FLAELME 199
    |||||
Db 371 FLAELME 377

RESULT 10
US-60-360-039-R006

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-991-150-8

Query Match      3.5%  Score 7; DB 5; Length 367;
Best Local Similarity 100.0%; Pred. No. 55;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 LAEKLE 181
    |||||
Db 356 LAEKLE 362

RESULT 6
US-10-137-337-685
; Sequence 685, Application US/10137337
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM28CIN
; CURRENT APPLICATION NUMBER: US/10/137,337
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: 09/758,447
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 812
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 685
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (66)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-137-337-685

```

```

Query Match      3.5%  Score 7; DB 6; Length 392;
Best Local Similarity 100.0%; Pred. No. 59;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 175 LAEKLE 181
    |||||
Db 111 LAEKLE 117

```

```

RESULT 7
US-60-360-039-17151
; Sequence 17151, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 4/374
; SEQ ID NO 17151
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-60-360-039-17151

```


GenCore version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

OM protein protein search, using sw model

Run on: July 2, 2002, 26.16.46, Search time 59.99 seconds
(without alignments)
338.459 Million cell updates/sec

Title: US-09-603-665-5_COPY_1917_2117
Perfect score: 201
Sequence: 1 FKLEDAKTERAPKDELLT.....NVILLFESIFELAEIEME 201

Scoring table: OLLGO
Gapop 60.0, Gapext 60.0

Searched: 307805 seqs, 10105545 residues

Word size: 6
Total number of hits satisfying chosen parameters: 405

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

- Database: Pending_Patents_AA_New**
1: /cqn2_6/ptodata/1/paa/pt_NEW_COMB pep:**
2: /cqn2_6/ptodata/1/paa/US05_NEW_COMB pep:**
3: /cqn2_6/ptodata/1/paa/US07_NEW_COMB pep:**
4: /cqn2_6/ptodata/1/paa/US08_NEW_COMB pep:**
5: /cqn2_6/ptodata/1/paa/US09_NEW_COMB pep:**
6: /cqn2_6/ptodata/1/paa/US10_NEW_COMB pep:**
7: /cqn2_6/ptodata/1/paa/US60_NEW_COMB pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	7	3.5	307	7	US-60-360-039-21343
2	7	3.5	317	7	US-60-360-039-6822
3	7	3.5	352	1	PRT-US02-00021-666
4	7	3.5	353	6	US-10-106-098-4568
5	7	3.5	367	5	US-09-991-150-8
6	7	3.5	392	6	US 10 137 337 685
7	7	3.5	393	7	US-60-360-039-17151
8	7	3.5	409	6	US-10-155-881-27684
9	7	3.5	415	7	US-60-360-039-2667
10	7	3.5	852	7	US-60-360-039-8096
11	7	3.5	893	7	US-60-360-039-9821
12	7	3.5	1045	6	US-10-041-018-199
13	7	3.5	1045	6	US-10-041-018-218
14	7	3.5	1045	6	US-10-041-018-264
15	7	3.5	1045	6	US-10-041-018-309
16	7	3.5	1045	6	US-10-041-018-336
17	7	3.5	1045	7	US-60-360-039-1836
18	7	3.5	1283	7	US-60-360-039-22781
19	7	3.5	1572	5	US-09-935-625-1959
20	7	3.5	1572	5	US-09-935-625-18337
21	7	3.5	1612	7	US 60 360 039 1628
22	7	3.5	1762	5	US-60-360-039-1628
23	7	3.5	1762	5	US-09-935-625-18336
24	7	3.5	1830	5	US-09-935-625-1957
25	7	3.5	1830	5	US-09-935-625-18335
26	7	3.5	2368	5	US-09-935-625-1963

ALIGNMENTS

RESULT 1
US-60-360-039-21343
; Sequence 21343, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; PUPPET FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21343
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Archaeoglobus fulgidus
US-60-360-039-21343

Query Match 3.5%; Score 7; DB 7; length 307;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 LAELKLE 181
DB 21 LAELKLE 27
|||||||
RESULT 2
US-60-360-039-6822
; Sequence 6822, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38 10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; PUPPET FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6822
; LENGTH: 317
; TYPE: PRT

US-09-758-451-174
; Sequence 174, Application US/09758451
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM018
; CURRENT APPLICATION NUMBER: US/09/758,451
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 316
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 174
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-451-174

Query Match 3.5% Score 7; DB 21; Length 48;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 74 LLLQFIL 80
|||||
Db 22 LLLQFIL 28

RESULT 15
US-09-621-976-4290
; Sequence 4290, Application US/09621976
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.H.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET:054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4290
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-4290

Query Match 3.5% Score 7; DB 20; Length 61;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 180 KPNYIVL 186
|||||
Db 30 KENYIVL 36

Search Completed: July 2, 2002, 16:14:39
Job Time: 696 sec

? PRIOR APPLICATION NUMBER: US 09/608,408
 ? PRIOR FILING DATE: 2000-06-30
 ? PRIOR APPLICATION NUMBER: US 09/774,203
 ? PRIOR FILING DATE: 2001-01-29
 ? NUMBER OF SEQ ID NOS: 49117
 ? SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ? SEQ ID NO 44493
 ? LENGTH: 46
 ? TYPE: PR1
 ? ORGANISM: Homo sapiens
 ? FEATURE:
 ? OTHER INFORMATION: MAP TO AC005229.1
 ? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 0.96
 ? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 0.9
 ? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL - 0.86
 ? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL - 1
 ? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL - 1.1
 ? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL - 0.86
 US-09-864 761 44493

Query Match 3.5% Score 7; DB 22; Length 36;
 Best local Similarity 100.0%; Pred. No. 76;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 GHEKEQ 122
 DB 4 GHEKEQ 10

RESULT 12

US-08-417-872A-12
 ? Sequence 32, Application US/08417872A
 ? GENERAL INFORMATION:
 ? APPLICANT: SPARKS, ANDREW H.
 ? APPLICANT: HOFFMAN, NOAH
 ? APPLICANT: KAY, BRIAN K.
 ? TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN
 ? TITLE OF INVENTION: OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME
 ? NUMBER OF SEQUENCES: 89
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Pennic & Edmonds LLP
 ? STREET: 1155 Avenue of the Americas
 ? CITY: New York
 ? STATE: New York
 ? COUNTRY: U.S.
 ? ZIP: 10036

? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent in Release #1.0, Version #1.25
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/417,872A
 ? FILING DATE: 07-APR-1995
 ? CLASSIFICATION: 536
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: MISTOCK, S. Leslie
 ? REGISTRATION NUMBER: 18,872
 ? REFERENCE/DOCKET NUMBER: 1101 193
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 212-790-9090
 ? TELEFAX: 212-864 9741/8864

? INFORMATION FOR SEQ ID NO: 32:

? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 41 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein
 ? HYPOTHETICAL: No
 ? ANTI-SENSE: No

? INFORMATION FOR SEQ ID NO: 32:

? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 41 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein
 ? HYPOTHETICAL: No
 ? ANTI-SENSE: No

US-08-417-872A-32

Query Match 3.5% Score 7; DB 8; Length 41;
 Best local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ENDPEKC 72
 DB 29 ENDPEKC 35

RESULT 13

US-09-879-957-79
 ? Sequence 79, Application US/09879957
 ? GENERAL INFORMATION:
 ? APPLICANT: SPARKS, ANDREW H.
 ? APPLICANT: HOFFMAN, NOAH
 ? APPLICANT: KAY, BRIAN K.
 ? APPLICANT: FOMKES, DENA M.
 ? APPLICANT: MCCONNELL, STEPHEN J.
 ? TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 ? TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 ? USING SAME
 ? NUMBER OF SEQUENCES: 227
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Pennic & Edmonds LLP
 ? STREET: 1155 Avenue of the Americas
 ? CITY: New York
 ? STATE: New York
 ? COUNTRY: USA
 ? ZIP: 10036-2711

? COMPUTER READABLE FORM:

? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent in Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/09/879,957
 ? FILING DATE: 13-Jun-2001
 ? CLASSIFICATION: Unknown
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 08/630,915
 ? FILING DATE: 03-APR-1996

? ATTORNEY/AGENT INFORMATION:

? NAME: MISTOCK, S. Leslie
 ? REGISTRATION NUMBER: 18,872
 ? REFERENCE/DOCKET NUMBER: 1101-174
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (212) 790-9090
 ? TELEFAX: (212) 864-8864/9741
 ? TELETYPE: 66141 PENNIE

? INFORMATION FOR SEQ ID NO: 79:

? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 41 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: Unknown
 ? TOPOLOGY: unknown
 ? MOLECULE TYPE: peptide
 ? SEQUENCE DESCRIPTION: SEQ ID NO: 79:

US-09-879-957-79

Query Match 3.5% Score 7; DB 22; Length 41;

Best local Similarity 100.0%; Pred. No. 75;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ENDPEKC 72
 DB 29 ENDPEKC 35

RESULT 14

```
; SEQ ID NO 6
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Chitinophaga pinensis
US-09-823-823-6

Query Match          4.0%; Score 8; DH 22; Length 220;
Best Local Similarity 100.0%; Pred No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 SENDPEKC 72
Db 130 SENDPEKC 137
|||||

RESULT 9
US-09-823-829-6
; Sequence 6; Application US/09823829
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Satoshi
; APPLICANT: Nakamura, Shoko
; APPLICANT: Suzuki, Makoto
; APPLICANT: Kasai, Hiroaki
; APPLICANT: Hamada, Tooru
; TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS
; FILE REFERENCE: 12817-004001
; CURRENT APPLICATION NUMBER: US/09/823,829
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/208,688
; PRIOR FILING DATE: 1998-12-10
; PRIOR APPLICATION NUMBER: JP 97/343316
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 6
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Chitinophaga pinensis
US-09-823-829-6

Query Match          4.0%; Score 8; DH 22; Length 220;
Best Local Similarity 100.0%; Pred No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 SENDPEKC 72
Db 130 SENDPEKC 137
|||||

RESULT 10
PCT-US01-006663-34446
; Sequence 34446; Application PCT/US01006663
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: PH 0004 W/ 7
; CURRENT APPLICATION NUMBER: PCT/US01/006663
; CURRENT FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04, 02, 00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 24 May 2000 (25, 05, 00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03, 08, 00)
; PRIOR APPLICATION NUMBER: GB 24263,6
```

```
; PRIOR FILING DATE: 03 October 2000 (03,10,00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27,09,00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21,09,00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30,06,00)
; NUMBER OF SEQ ID NOS: 38837
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 34446
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005229.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
PCT-US01-006663-34446

Query Match          3.5%; Score 7; DH 1; Length 36;
Best Local Similarity 100.0%; Pred No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 GCEKFKQ 122
Db 4 GCEKFKQ 10
|||||

RESULT 11
US-09-864-761-44493
; Sequence 44493; Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aesmlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
```

PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/577,409
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: PCT/US01/04927
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 09/519,705
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: US 09/574,454
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: PCT/US01/04941
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: US 09/540,217
PRIOR FILING DATE: 2000-04-31
PRIOR APPLICATION NUMBER: US 09/649,167
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-04-30
PRIOR APPLICATION NUMBER: US 09/552,929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/770,160
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/US01/08656
NUMBER OF SEQ ID NOS: 412
SOFTWARE: pF_LGenes Version 6.0
SEQ ID NO 167
LENGTH: 1149
TYPE: PRT
ORGANISM: Homo sapiens
US: 60-339,454,167

Query Match 49.8% Score 100; DB 26; Length 1149;
Best Local Similarity 100.0%; Pred. No. 9,4e-94;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ALMPPLVQLENRIIGGEKFKQVHTKHLPCIAQFSVAMADSLWKPLNYQILLKTRDSS 161
|||||
DB 1023 ALMPPLVQLENRIIGGEKFKQVHTKHLPCIAQFSVAMADSLWKPLNYQILLKTRDSS 1082

QY 162 PKVFEALITVLALAEKLENIVLLPESIPFLAEIMEDE 201
|||||
DB 1084 PKVFEALITVLALAEKLENIVLLPESIPFLAEIMEDE 1122

RESULT 7
US 60 141 423 4
Sequence 3, Application US/60141423
GENERAL INFORMATION:
APPLICANT: Barry, Caroline
APPLICANT: Rouqueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: A NOVEL HAP28 GENE AND PROTEIN
FILE REFERENCE: GENSET 0649F
CURRENT APPLICATION NUMBER: US/60/141,424
CURRENT FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent.com
SEQ ID NO 3
LENGTH: 2144
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 91...111
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 212...242
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 700...720

OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 861...881
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 884...904
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 911...931
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 937...957
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 1297...1317
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 1410...1430
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: 1725...1745
OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
FEATURE:
NAME/KEY: SITE
LOCATION: 1750...1771
OTHER INFORMATION: potential leucine zipper pattern, BLA Program
US: 60-141-323-3

Query Match 49.8% Score 100; DB 26; Length 2144;
Best Local Similarity 100.0%; Pred. No. 1.7e-93;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ALMPPLVQLENRIIGGEKFKQVHTKHLPCIAQFSVAMADSLWKPLNYQILLKTRDSS 161
|||||
DB 2018 ALMPPLVQLENRIIGGEKFKQVHTKHLPCIAQFSVAMADSLWKPLNYQILLKTRDSS 2077

QY 162 PKVFEALITVLALAEKLENIVLLPESIPFLAEIMEDE 201
|||||
DB 2078 PKVFEALITVLALAEKLENIVLLPESIPFLAEIMEDE 2117

RESULT 8
US: 09-823-823-6
Sequence 6, Application US/09823823
GENERAL INFORMATION:
APPLICANT: Yamamoto, Satoshi
APPLICANT: Kasai, Hiroaki
APPLICANT: Nakamura, Shoko
APPLICANT: Suzuki, Makoto
APPLICANT: Hamada, Tohru
TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS USING
FILE REFERENCE: 12817-004001
CURRENT APPLICATION NUMBER: US/09/823,823
CURRENT FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 09/208,688
PRIOR FILING DATE: 1998-12-10
PRIOR APPLICATION NUMBER: JP 97/343316
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 2.0

QY 162 PKVRFALITVLAELKLNKENVIVLLPESIPFLAELMEDE 201
 DB 449 PKVRFALITVLAELKLNKENVIVLLPESIPFLAELMEDE 488

RESULT 4

US-60-042-611-52
 ? Sequence 52, Application US/60042611
 ? GENERAL INFORMATION:
 ? APPLICANT: Howcock, Anne M.
 ? TITLE OF INVENTION: Compositions and Methods Comprising
 ? TITLE OF INVENTION: BARD1 and Other BRCA1 Binding Proteins
 ? NUMBER OF SEQUENCES: 115
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Arnold, White & Durkee
 ? STREET: P.O. Box 4433
 ? CITY: Houston
 ? STATE: Texas
 ? COUNTRY: USA
 ? ZIP: 77210
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: PatentIn Release #1.0, Version #1.30
 ? CURRENT APPLICATION NUMBER: US/60/042-611
 ? FILING DATE:
 ? CLASSIFICATION:
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Hibler, David W.
 ? REGISTRATION NUMBER: P-41,071
 ? REFERENCE/DOCKET NUMBER: UTSD:499PZ2
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 512/418-3000
 ? TELEFAX: 512/474-7577
 ? INFORMATION FOR SEQ ID NO: 52:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 515 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS:
 ? TOPOLOGY: linear
 ? US-60-042-611-52

Query Match 49.8%, Score 100, DB 26, Length 515;
 Best Local Similarity 100.0%; Pred. No. 4.5e-94;
 Matches 100; Conservative 0; Mismatches 0; Indels 0, Gaps 0;
 QY 102 ALMPLVDLENRLGGKFKFQVRVTKHLIPCIAGFSVAMADDSLWKPLNYQILLKTRDSS 161
 DB 389 ALMPLVDLENRLGGKFKFQVRVTKHLIPCIAGFSVAMADDSLWKPLNYQILLKTRDSS 448

RESULT 5

US-60-042-985-52
 ? Sequence 52, Application US/60042985
 ? GENERAL INFORMATION:
 ? APPLICANT: Howcock, Anne M.
 ? TITLE OF INVENTION: Compositions and Methods Comprising
 ? TITLE OF INVENTION: BARD1 and Other BRCA1 Binding Proteins
 ? NUMBER OF SEQUENCES: 115
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Arnold, White & Durkee
 ? STREET: P.O. Box 4433
 ? CITY: Houston
 ? STATE: Texas

COUNTRY: USA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION NUMBER: US/60/042-985
 FILING DATE: Concurrently Herewith
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Hibler, David W.
 REGISTRATION NUMBER: P-41,071
 REFERENCE/DOCKET NUMBER: UTSD:499PZ3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 512/418-3000
 TELEFAX: 512/474-7577
 INFORMATION FOR SEQ ID NO: 52:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 515 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-60-042-985-52

Query Match 49.8%, Score 100, DB 26, Length 515;
 Best Local Similarity 100.0%; Pred. No. 4.5e-94;
 Matches 100; Conservative 0; Mismatches 0; Indels 0, Gaps 0;
 QY 102 ALMPLVDLENRLGGKFKFQVRVTKHLIPCIAGFSVAMADDSLWKPLNYQILLKTRDSS 161
 DB 389 ALMPLVDLENRLGGKFKFQVRVTKHLIPCIAGFSVAMADDSLWKPLNYQILLKTRDSS 448

RESULT 6

US-60-339-453-167
 ? Sequence 167, Application US/60339453
 ? GENERAL INFORMATION:
 ? APPLICANT: Tang, Y. Tom
 ? APPLICANT: Wang, Zhiwei
 ? APPLICANT: Wang, Jian-Rui
 ? APPLICANT: Ghosh, Malabika
 ? APPLICANT: Wenq, Gezhi
 ? APPLICANT: Boyle, Bryan J
 ? APPLICANT: Drmanac, Radoje T
 ? TITLE OF INVENTION: Novel Nucleic Acids and
 ? TITLE OF INVENTION: Polypeptides
 ? FILE REFERENCE: 812
 ? CURRENT APPLICATION NUMBER: US/60/339,453
 ? PRIOR FILING DATE: 2003-03-12
 ? PRIOR APPLICATION NUMBER: US 09/488,725
 ? PRIOR FILING DATE: 2000-01-21
 ? PRIOR APPLICATION NUMBER: US 09/552,317
 ? PRIOR FILING DATE: 2000-04-25
 ? PRIOR APPLICATION NUMBER: PCT/US00/35017
 ? PRIOR FILING DATE: 2000-12-22
 ? PRIOR APPLICATION NUMBER: US 09/491,404
 ? PRIOR FILING DATE: 2000-01-25
 ? PRIOR APPLICATION NUMBER: PCT/US01/02623
 ? PRIOR FILING DATE: 2001-01-25
 ? PRIOR APPLICATION NUMBER: US 09/496,914
 ? PRIOR FILING DATE: 2000-02-03
 ? PRIOR APPLICATION NUMBER: US 09/560,875
 ? PRIOR FILING DATE: 2000-04-27
 ? PRIOR APPLICATION NUMBER: PCT/US01/03800
 ? PRIOR FILING DATE: 2001-02-05
 ? PRIOR APPLICATION NUMBER: US 09/515,126

```

; LOCATION: 1967
; OTHER INFORMATION: Xaa-Asp or Asn
; NAME/KEY: VARIANT
; LOCATION: 2017
; OTHER INFORMATION: Xaa-Gly or Glu
US 09-603-665-5

Query Match          99.0%  Score 199; DB 20; Length 2144.
Best Local Similarity 100.0%; Pred. No. 6,20-196;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKLFWAKTEGAPKRLITFYNLADGIAEKLGKGLFTLFAGHLVKPFAALITLVLAIAEKIK 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1917 FKLFWAKTEGAPKRLITFYNLADGIAEKLGKGLFTLFAGHLVKPFAALITLVLAIAEKIK 1976

QY 61 AFDSNDPEKCCLLQFLLNCLYKIFLFTQHFISKPRAXALMPLVQLENPLGGEEK 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1977 AFDSNDPEKCCLLQFLLNCLYKIFLFTQHFISKPRAXALMPLVQLENPLGGEEK 2036

QY 121 FQPVTKHLIPCIAPFSVAMADSLWKPINYOILKTRDSSPKVRFAALITVLALAEKIK 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 2037 FQPVTKHLIPCIAPFSVAMADSLWKPINYOILKTRDSSPKVRFAALITVLALAEKIK 2096

QY 181 ENVIVLLPESIPFLAEIMEDE 201
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 2097 ENVIVLLPESIPFLAEIMEDE 2117

RESULT 3
US-08-936-487-52
; Sequence 52, Application US/08046487
; GENERAL INFORMATION:
; APPLICANT: Bowcock, Anne M.
; APPLICANT: Hager, Richard
; TITLE OF INVENTION: Compositions and Methods Comprising
; TITLE OF INVENTION: BAK1 and Other BRCA1 Binding Proteins
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,487
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 60/045,296
; FILING DATE: 20-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,611
; FILING DATE: 03-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,985
; FILING DATE: 04-APR-1997
; ATTORNEY/AGENT INFORMATION
; NAME: Hibler, David W.
; REGISTRATION NUMBER: p-61,071
; REFERENCE/DOCKET NUMBER: UTSD:499
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-936-487-52

Query Match          49.8%  Score 100; DB 13; Length 515;
Best Local Similarity 100.0%; Pred. No. 4,5e-94;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ALMMPLVQLENPLGGEEKFQKVKHLIPCIAPFSVAMADSLWKPINYOILKTRDSS 161
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 389 ALMMPLVQLENPLGGEEKFQKVKHLIPCIAPFSVAMADSLWKPINYOILKTRDSS 448

; LOCATION: 1967
; OTHER INFORMATION: Xaa-Asp or Asn
; NAME/KEY: VARIANT
; LOCATION: 2017
; OTHER INFORMATION: Xaa-Gly or Glu
US 09-603-665-5

Query Match          99.0%  Score 199; DB 20; Length 2144.
Best Local Similarity 100.0%; Pred. No. 6,20-196;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKLFWAKTEGAPKRLITFYNLADGIAEKLGKGLFTLFAGHLVKPFAALITLVLAIAEKIK 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1917 FKLFWAKTEGAPKRLITFYNLADGIAEKLGKGLFTLFAGHLVKPFAALITLVLAIAEKIK 1976

QY 61 AFDSNDPEKCCLLQFLLNCLYKIFLFTQHFISKPRAXALMPLVQLENPLGGEEK 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1977 AFDSNDPEKCCLLQFLLNCLYKIFLFTQHFISKPRAXALMPLVQLENPLGGEEK 2036

QY 121 FQPVTKHLIPCIAPFSVAMADSLWKPINYOILKTRDSSPKVRFAALITVLALAEKIK 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 2037 FQPVTKHLIPCIAPFSVAMADSLWKPINYOILKTRDSSPKVRFAALITVLALAEKIK 2096

QY 181 ENVIVLLPESIPFLAEIMEDE 201
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 2097 ENVIVLLPESIPFLAEIMEDE 2117

RESULT 2
US-60-176-880-4
; Sequence 4, Application US/60176880
; GENERAL INFORMATION:
; APPLICANT: Harry, Caroline
; APPLICANT: Houqueleret, Lydia
; APPLICANT: Choumakov, Ilya
; TITLE OF INVENTION: A NOVEL BAP28 GENE AND PROTEIN.
; FILE REFERENCE: 67 US2 PPO
; CURRENT APPLICATION NUMBER: US/60/176,880
; CURRENT FILING DATE: 2000-01-18
; EARLIER APPLICATION NUMBER: US 60/141,423
; EARLIER FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent.pm
; SEQ ID NO 4
; LENGTH: 2144
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 1694
; OTHER INFORMATION: Xaa-Ser or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 1854
; OTHER INFORMATION: Xaa-Ala or Val
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 1967
; OTHER INFORMATION: Xaa-Asp or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 2017
; OTHER INFORMATION: Xaa-Gly or Glu
US 60-176-880-4

Query Match          99.0%  Score 199; DB 20; Length 2144.
Best Local Similarity 100.0%; Pred. No. 6,20-196;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKLFWAKTEGAPKRLITFYNLADGIAEKLGKGLFTLFAGHLVKPFAALITLVLAIAEKIK 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1917 FKLFWAKTEGAPKRLITFYNLADGIAEKLGKGLFTLFAGHLVKPFAALITLVLAIAEKIK 1976
```

GenGene version 4.5			
Copyright (c) 1993 - 2000 CompuGen Ltd			
M protein - protein search, using sw model			
Run on	July 2, 2002, 16:14:37	Search time: 506.3 seconds (without alignments)	140,012 Million cell updates/sec
Title:	US-09-603-665-5_COPY_1917_2117		
Perfect score:	261		
Sequence:	1 PKLEFWAKTDAAPKQKLLP.....NVLVLLPSPFLALMLME 201		
Scoring table:	OLIGO		
	Gapop 60 0 0	Gapext 60 0 0	
Searched:	3502263 seqs, 303480561 residues		
Word size:	6		
Total number of hits satisfying chosen parameters:	2344		
Minimum DB seq length:	0		
Maximum DB seq length:	2600000000		
Post-processing:	Listing first 45 summaries		
Database:	Pending patents, AA, Main:		
	1. /cgn2_6/ptodata/2/paa/US06_COMB pep.*		
	2. /cgn2_6/ptodata/2/paa/US06_COMB pep.*		
	3. /cgn2_6/ptodata/2/paa/US07_COMB pep.*		
	4. /cgn2_6/ptodata/2/paa/US08_COMB pep.*		
	5. /cgn2_6/ptodata/2/paa/US081_COMB pep.*		
	6. /cgn2_6/ptodata/2/paa/US082_COMB pep.*		
	7. /cgn2_6/ptodata/2/paa/US083_COMB pep.*		
	8. /cgn2_6/ptodata/2/paa/US084_COMB pep.*		
	9. /cgn2_6/ptodata/2/paa/US085_COMB pep.*		
	10. /cgn2_6/ptodata/2/paa/US086_COMB pep.*		
	11. /cgn2_6/ptodata/2/paa/US087_COMB pep.*		
	12. /cgn2_6/ptodata/2/paa/US088_COMB pep.*		
	13. /cgn2_6/ptodata/2/paa/US089_COMB pep.*		
	14. /cgn2_6/ptodata/2/paa/US090_COMB pep.*		
	15. /cgn2_6/ptodata/2/paa/US091_COMB pep.*		
	16. /cgn2_6/ptodata/2/paa/US092_COMB pep.*		
	17. /cgn2_6/ptodata/2/paa/US093_COMB pep.*		
	18. /cgn2_6/ptodata/2/paa/US094_COMB pep.*		
	19. /cgn2_6/ptodata/2/paa/US095_COMB pep.*		
	20. /cgn2_6/ptodata/2/paa/US096_COMB pep.*		
	21. /cgn2_6/ptodata/2/paa/US097_COMB pep.*		
	22. /cgn2_6/ptodata/2/paa/US098_COMB pep.*		
	23. /cgn2_6/ptodata/2/paa/US099_COMB pep.*		
	24. /cgn2_6/ptodata/2/paa/US100_COMB pep.*		
	25. /cgn2_6/ptodata/2/paa/US101_COMB pep.*		
	26. /cgn2_6/ptodata/2/paa/US60_COMB pep.*		
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	199	99 0 2144 20 US 09 603 665 5	Sequence 5, Appli
2	199	99 0 2144 26 US-60-176 880 4	Sequence 4, Appli
3	100	49 8 515 14 US-08-946-487-52	Sequence 52, Appl
4	100	49 8 515 26 US-60-141-323	Sequence 52, Appl
5	100	49 8 515 26 US-60-642-985 52	Sequence 52, Appl
6	100	49 8 1149 26 US-60-339 453 167	Sequence 167, App
7	100	49 8 2144 26 US 60 141 323 3	Sequence 3, Appli

Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli	Sequence 6, Appli
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

ALIGNMENTS			
RESULT 1			
US-09-603-665-5	220	4.0	Sequence 6, Appli
Sequence 5, Application US/0903665	220	4.0	Sequence 6, Appli
GENERAL INFORMATION:	220	4.0	Sequence 6, Appli
APPLICANT: Buitz, Caroline	220	4.0	Sequence 6, Appli
APPLICANT: Boogachet, Lydie	220	4.0	Sequence 6, Appli
APPLICANT: Chumakov, Liya	220	4.0	Sequence 6, Appli
APPLICANT: Cohen, Akevine, Annick	220	4.0	Sequence 6, Appli
TITLE OF INVENTION: A NOVEL BAF28 GENE AND PROTEIN	220	4.0	Sequence 6, Appli
FILE REFERENCE: GENESET-063ADS	220	4.0	Sequence 6, Appli
CURRENT APPLICATION NUMBER: US/09/003,665	220	4.0	Sequence 6, Appli
PRIOR FILING DATE: 2000-06-23	220	4.0	Sequence 6, Appli
PRIOR APPLICATION NUMBER: US 60/141,323	220	4.0	Sequence 6, Appli
PRIOR FILING DATE: 1999-06-25	220	4.0	Sequence 6, Appli
PRIOR APPLICATION NUMBER: US 60/176,880	220	4.0	Sequence 6, Appli
PRIOR FILING DATE: 2000-01-18	220	4.0	Sequence 6, Appli
NUMBER OF SEQ ID NOS: 63	220	4.0	Sequence 6, Appli
SOFTWARE: Patent.pm	220	4.0	Sequence 6, Appli
SEQ ID NO 5	220	4.0	Sequence 6, Appli
LENGTH: 2144	220	4.0	Sequence 6, Appli
TYPE: PRT	220	4.0	Sequence 6, Appli
ORGANISM: Homo sapiens	220	4.0	Sequence 6, Appli
FEATURE:	220	4.0	Sequence 6, Appli
NAME/KEY: VARIANT	220	4.0	Sequence 6, Appli
LOCATION: 1694	220	4.0	Sequence 6, Appli
OTHER INFORMATION: Xaa-Ser or Asn	220	4.0	Sequence 6, Appli
NAME/KEY: VARIANT	220	4.0	Sequence 6, Appli
LOCATION: 1854	220	4.0	Sequence 6, Appli
OTHER INFORMATION: Xaa-Ala or Val	220	4.0	Sequence 6, Appli
NAME/KEY: VARIANT	220	4.0	Sequence 6, Appli

Query Match 3.5% Score 7; DB 22; Length 289;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAA 10
 |||||
 Db 73 LLSALAA 79

Search completed: July 2, 2002, 16:06:03
 Job time: 180 sec

Query Match 3.5% Score 7; DB 22; Length 254;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 ENDLEEV 136
 |||||
 Db 15 endleev 21

RESULT 15

ABG01568
 ID ABG01568 standard; Protein; 289 AA.

XX AC AHC01568;

XX DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #1559.

XX Human; chromosome mapping, gene mapping; gene therapy, forensic;
 food supplement; medical imaging, diagnostic, genetic disorder.

XX OS Homo sapiens.

XX PN W0200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-0508631.

XX PR 31-MAR-2000; 2000US-0540217

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX PR WPI: 2001-639362/73

XX DR N-PSDH; AAS65755.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX PS Claim 20; SEQ ID No 31927; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG0017 represent novel human
 CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX CC Sequence 289 AA:

50

Best Local Similarity 100.0%; Score 7; PB 22; Length 245;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AAMAQK 13
II 119 aaaaqk 125
II 119 aaaaqk 125

RESULT 14
AA035868
ID AA035868 standard; Protein: 254 AA.
XX
AC AA035868;
XX
DT 14-FEB-2002 (first entry)
XX
DE Helicobacter pylori cellular proliferation protein #181.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
CS Helicobacter pylori.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PY 21-MAR 2001; 2001WO-US0918).
XX
PR 21-MAR 2000; 2000US-191078).
PR 23-MAY-2000; 2000US-206848).
PR 26-MAY-2000; 2000US-207727).
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22 DEC 2000; 2000US-257931P.
PR 15 FEB 2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlson KL, Syskind JW, Wall G, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR N-PSDB; AAS53727.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; Seq ID No 11461, 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO et
XX http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 254 AA;

Best Local Similarity 100.0%; Pred. No. 1,26-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AAMAQK 13
II 119 aaaaqk 125
II 119 aaaaqk 125

RESULT 14
AA095242
ID AA095242 standard; Protein: 245 AA.
XX
AC AA095242;
XX
DT 26 JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO-17469
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
CS Homo Sapiens.
XX
PN EP1074617 A2.
XX
PD 07 FEB-2001.
XX
PY 28 JUL-2000; 2000EP-0116126.
XX
PR 29 JUL 1999; 95JP 0248036.
PR 27 AUG-1999; 95JP 0-00254.
PR 11 JAN-2000; 2000JP 0118776.
PR 02 MAY-2000; 2000JP 0343727.
PR 09 JUN-2000; 2000JP 0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isonai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama I, Wakamatsu A, Nishii K, Otsuki T;
XX
XX WPI: 2001-08749/4.
DR
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full length cDNAs
XX
PS Claim 8; SEQ ID 17469; 2537pp - CD ROM; English.
XX
CC the present invention describes primer sets for synthesizing 5602
CC full length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAR0366 to AAR13629 and
CC AAR13633 to AAR19742 represent human cDNA sequences; AAR02446 to
CC AAR95894 represent human amino acid sequences; and AAR13629 to AAR13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 245 AA;

OS Escherichia coli.
 XX JP05091885-A.
 PN 16-APR-1993.
 PD 06-SEP-1991; 91JP-0254341.
 PF 06-SEP-1991; 91JP-0254341.
 XX 06-SEP-1991; 91JP-0254341.
 PR (NIGE-) NIPPON GENE KK.
 XX WPI: 1993-161735/20.
 XX N-PSDB; AAQ42389.
 DR Gene cloning vector for easy selection, comprises regulatory
 XX gene with coding aminoacid sequence, and contg. numerous
 PT restriction enzyme sites, avoiding specific selection medium or
 PT chromogen for lac Z
 XX Disclosure; Page 7; 17pp; Japanese.
 PS The sequence is that of an rcsB related gene prod. having the same
 XX amino acid sequence as the rcsB gene (the cps regulatory gene) and
 CC contg. restriction enzyme sites. The gene is carried on the
 CC cloning vector PERISH7-alpha, in which the rcsB related gene is rcsB-
 CC alpha which contains 7 substituted bases and the ampicillin
 CC resistance gene from pUC18. Using this new cloning vector the
 CC specific selection medium or chromogen for lac Z is not required.
 CC See also AAR36606.
 XX Sequence 216 AA;
 SQ

Query Match 3.5%; Score 7; DB 14; Length 216;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALAALQK 13
 DB 119 alaalgk 125
 |||||

RESULT 11
 ID AAR36606 standard; Protein; 216 AA.
 XX AC AAR36606;
 XX DT 26-AUG-1993 (first entry)
 DE RcsB related gene prod. from PERISH7-alpha.
 XX Cps regulatory gene; restriction enzyme site; lac Z; chromogen.
 XX Escherichia coli.
 XX JP05091885-A.
 XX 16-APR-1993.
 XX 06-SEP-1991; 91JP-0254341.
 XX 06-SEP-1991; 91JP-0254341.
 XX (NIGE-) NIPPON GENE KK.
 XX WPI: 1993-161735/20.
 XX N-PSDB; AAQ42390.
 DR Gene cloning vector for easy selection, comprises regulatory
 XX gene with coding aminoacid sequence, and contg. numerous
 PT restriction enzyme sites, avoiding specific selection medium or
 PT

PT chromogen for lac Z
 XX Disclosure; Page 8; 17pp; Japanese.
 XX The sequence is that of an rcsB related gene prod. having the same
 CC amino acid sequence as the rcsB gene prod. (the cps regulatory gene)
 CC and contg. restriction enzyme sites. The gene is carried on the
 CC cloning vector PERISH7-alpha, in which the rcsB related gene is rcsB-
 CC alpha which contains 7 substituted bases and the ampicillin
 CC resistance gene from pUC18 (AATACT is substd. for tch Seal
 CC recognition site AclACT). Using this new cloning vector the
 CC specific selection medium or chromogen for lac Z is not required.
 CC See also AAR36605.
 XX Sequence 216 AA;
 SQ

Query Match 3.5%; Score 7; DB 14; Length 216;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALAALQK 13
 DB 119 alaalgk 125
 |||||

RESULT 12
 ID AAR31999 standard; Protein; 216 AA.
 XX AC AAR31999;
 XX DT 03-JUN-1993 (first entry)
 DE Prod. of the rcsB gene
 XX E. coli; capsulat polysaccharide, synthetic gene; PERISH1;
 KW selection; regulation
 XX Escherichia coli.
 XX JP04365485-A.
 XX 17-DEC-1992.
 XX 02-JUL-1991; 91JP-0186956.
 XX 03-AUG-1990; 90JP-0205160.
 XX (APAK/) ARAKAWA N.
 XX WPI: 1993-040507/05.
 XX N-PSDB; AAQ34548.
 XX Selection of foreign DNA fragment recombinants - using the rcsB
 PT gene in a vector, into which is inserted the foreign DNA at a
 PT specified restriction enzyme site, and using the vector to
 PT transform E.coli HB101 cells
 XX Disclosure; Page 5; 7pp; Japanese.
 XX The sequence is produced by the rcsB gene region, the regulatory gene
 CC of E. coli capsular polysaccharide synthetic gene. The gene may be
 CC inserted in a vector and foreign DNA inserted at the rcsB gene
 CC locus. The ligated vector (e.g. PERISH1) may be used to transform
 CC E. coli HB101 cells, with only recombinant cells being cultured.
 CC The gene provides a method for selection of recombinant clones,
 CC without the need for replica or colour media.
 XX Sequence 216 AA;
 SQ

Query Match 3.5%; Score 7; DB 14; Length 216;

SQ Sequence 195 AA:

Query Match 4.58; Score 7; DH 22; Length 195;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LAAALQKV 14
| | | | |
Db 45 LAAALQKV 41

RESULT 8

AAK13165

ID AAK13165 standard; Protein: 200 AA.

XX AAK13165;

XX AAK13165;

DT 09-OCT-1991 (first entry)

DE Eimeria merozoite surface antigen precursor.

XX Merozoite; antigen; vaccine; coccidiosis.

OS Eimeria tenella.

XX Key Location/Qualifiers

PI Peptide 1..20

FI /Label= sig_peptide

FT Protein 21..210

FI /Label= mat_peptide

XX EP439056 A.

XX 41 JUL 1991.

XX 17 JAN 1991; 91EP 0100502.

XX 26 JAN 1990; 90US 0470508.

XX (HOFF) HOFFMANN LA ROCHE AC.

XX Biofer MB;

XX WPI; 1991-224487/11.

DR N-PSDB; AAC12782.

XX Eimeria merozoite surface antigen (2.4kb), cDNA encoding it and

recombinant virus, useful in the prodn. of vaccine against

coccidiosis in poultry

XX Disclosure: Fig 1(a b); 40pp; English.

XX The precursor is believed to be processed by proteolysis in vivo to

the Eimeria surface antigen.

XX A recombinant vector of virus contg. the DNA sequence encoding this

protein is used in a vaccine for protecting poultry against coccidiosis,

i.e. infection by E. tenella, E. acervulina and E. maxima.

XX The dosage is 5-50 microg/kg, pref. 25 microg/kg, followed by booster

XX vaccinations.

Query Match

Best Local Similarity 4.58; Score 7; DH 12; Length 200;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LLSALAA 10
| | | | |
Db 56 LLSALAA 62

RESULT 9

ABB69008

ID ABB69008 standard; Protein: 200 AA.

XX ABB69008;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 33816.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-080231.

XX 23-MAR-2000; 2000US-1916379.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PND, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL13111.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signalling and cell-cell

interactions.

XX Disclosure; SEQ ID NO 33816; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from Drosophila. The invention is

useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of

insecticides, therapeutics and pharmaceutical drugs. The invention

discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

sequences (ABL01840-ABL16175) and the encoded proteins

(AHE57737 AHE72072).

XX The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 200 AA;

Query Match

Best Local Similarity 3.58; Score 7; DH 22; Length 200;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LLSALAA 10
| | | | |
Db 85 LLSALAA 91

RESULT 10

AAK36605

ID AAK36605 standard; Protein: 216 AA.

XX AAK36605;

XX 26-AUG-1993 (first entry)

XX KcsH related gene prod. from pKISH6-alpha.

XX Gps regulatory gene; restriction enzyme site; lac Z; chromogen.

CC polynucleotide (PCR) primers, oligomers, and for chromosomal
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).
XX
SQ Sequence 98 AA:

Query Match 3.5%; Score 7; DB 22; Length 98;
Best Local Similarity 180.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 TILAPRV 68
| | | | |
DB 84 ttilaprv 90

RESULT 6
AAB27313
ID AAB27313 standard; Protein: 165 AA.

AC AAB27313;

XX 25-JAN-2001 (first entry)

DE B. vulgaris NIM1 homologue SEQ ID NO: 40.

XX Systemic acquired resistance, SAR, signal transduction cascade;
KW disease resistance: tobacco, tomato, chili, sunflower, sugarbeet,
KW potato.

XX Beta vulgaris.

XX WO200053762-A2.

PN 14-SEP-2000.

PD 07-MAR-2000, 2000W0-EF0147R

FF N-PSDB; AAB27216.

XX 09-MAR-1999; 99US-0265149.

XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-EFFINDINGEN VERW GES MBH.

XX Salmeron JM, Weisio LJ, Willis MG, Mengiste T;

XX WPI: 2000-594322/56.

DR N-PSDB; AAB27216.

XX Novel plant genes for enhancing systemic acquired resistance gene

PT expression and broad spectrum disease resistance in plants, are

PT homologues of Arabidopsis NIM1 gene

XX Claim 1: Page 111; 152pp; English

XX The present invention is concerned with the isolation of NIM1 homologues

CC and their coding sequences from Arabidopsis thaliana, Brassica napus,
CC Nicotiana tabacum, Lycopersicon esculentum, Beta vulgaris, Helianthus

CC annuus and Solanum tuberosum. NIM1 is one of the proteins involved in the
CC signal transduction cascade leading to systemic acquired resistance (SAR)
CC in plants. This gives the plants an increased resistance to disease. The
CC protein and gene can be used to produce transgenic plants resistant to
CC diseases caused by viruses, such as tobacco etch virus, red clover mottle virus,
CC ringspot virus, pelargonium leaf curl virus, red clover mottle virus,
CC tomato bushy stunt virus, fungi, including Phytophthora parasitica and
CC Puccinia tabacina, bacteria such as Pseudomonas syringae and P.
CC tabaci, insects, including aphids and lepidoptera and nematodes such as
CC Meloidogyne incognita. In particular they can be used against disease
CC organisms of maize.
XX
SQ Sequence 165 AA;

Query Match 3.5%; Score 7; DB 21; Length 165;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 DAPKDLR 180
| | | | |
DB 115 dapkdlr 121

RESULT 7
AAB63428
ID AAB63428 standard; Protein: 195 AA.

XX AAB63428;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 17076.

XX Drosophila, developmental biology, cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001W0-US09231.

XX 23-MAR-2000; 2000US-191637P.

PP 11-MAR-2000; 2000US-0614150.

XX (PPEK) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-556860/75.

DR N-PSDB; ABL07531.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure: SEQ ID NO 17076; 21pp; Sequence listing, English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB16181), expressed DNA
CC sequences (AB161840-AB16175) and the encoded proteins
CC (AB257737-AB257202).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).
XX

```

Q10TEZ
ID Q10TEZ PRELIMINARY PRI 98 AA
AC Q10TEZ
DI 01 JUN 2001 (11EMBLrel, 17, Created)
DI 01 JUN 2001 (11EMBLrel, 17, Last sequence update)
DI 01 DEC 2001 (11EMBLrel, 19, Last annotation update)
DE SUCCINATE-COENZYME A LIGASE, ADP-FORMING, BETA SUBUNIT
DE (FRAGMENT).
DE
EN
GN Mus musculus (Mouse).
OS
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NBL_TaxID=10090;
EN
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=EMHRYO;
FX MEDLINE=21085660; PubMed=11217851;
KA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
KA Arakawa T., Hara A., Fukunishi Y., Kanno H., Adachi J., Fukuda S.,
KA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
KA Saito T., Okazaki Y., Gojobori I., Hono H., Kasukawa T., Saito R.,
KA Kadota K., Matsuda H.A., Ashburner M., Hatalov S., Casavant J.,
KA Fleischmann W., Gausterland J., Gliss C., King B., Kouchiwa H.,
KA Kuchi P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
KA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
KA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
KA Hlake J., Hoffell D., Hojunga N., Carninci P., de Bonaldo M.F.,
KA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
KA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
KA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
KA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
KA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
KA Suzuki H., Toyooka K., Water K.H., Weitz C., Whitaker C., Wilming L.,
KA Wyshaw Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
KA Hayashizaki Y.
RT *Functional annotation of a full-length mouse cDNA collection.*;
PL Nature 409:685-690(2001).
DB EMBL: AK004867; BAB2948.1;
DB HSSP: P07460; ZSR9.
DB MGI: MGI:130677; Sucta2.
FT NON-TER 1
SQ SEQUENCE 98 AA: 1041 MW: 95E0167C6A2D9F07 CEG64;

```

```

Query Match 3.5% Score 7; DB 11; Length 98;
Best local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
9Y 47 MVVKLSE 43
DB 74 MVVKLSE 80

```

Search completed: July 2, 2002, 16:18:31
 Job time: 86.3 sec

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission";
RU Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: 055857; AAA98035.2; -;
DR EMBL: 070855; AAB09160.2; -;
SQ SEQUENCE 895 AA; 103130 MW; 55AFD/041F235037 CRC64;

Query Match 4.0%; Score 8; DB 5; Length 895;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 NDLEEVG 24
| | | | | | | |
DB 696 NDLEEVG 703

RESULT 12
Q9X617 PRELIMINARY; PRT; 56 AA.

AC Q9X617;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ALANYL-TRNA SYNTHETASE (FRAGMENT).
GN ALAS.

OS Flavobacterium ferrugineum.
OC Bacteria; CFb group; Flavobacteriia; Flavobacteriaceae; Flavobacterium.
OX NCBI_TaxID=249;
[1]

RP SEQUENCE FROM N.A.
RC STRAIN-ATCC13524;
RA Gupta R.S.; Mukhtar T.; Singh B.;
RX MEDLINE:99058143; PubMed:9841678;

RA Gupta R.S.;
RT "Protein phylogenies and signature sequences: A reappraisal of
RT evolutionary relationships among archaeobacteria, eubacteria, and
RT eukaryotes.";
PL Microbiol. Mol. Biol. Rev. 62:1435-1491(1998).

FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC13524;

RA Gupta R.S.; Mukhtar T.; Singh B.;
RT "Evolutionary relationships among photosynthetic prokaryotes
RT (Hellobacterium chlorum, Chloroflexus auranticus, Cyanobacteria,
RT Chlorobium tepidum and Proteobacteria): implications regarding the
RT origin of photosynthesis.";
PL Mol. Microbiol. 0:0-0(1999).

DR EMBL: AFI30451; AAD33469.1;
DR InterPro: IPR002318; trna-synt_2c;
DR Pfam: PF01411; trna-synt_2c; 1;
KW Aminoacyl-tRNA synthetase.

FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6451 MW; 8623CE781D1B4E7 CRC64;

Query Match 3.5%; Score 7; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 NDLEEVG 24
| | | | | | | |
DB 38 NDLEEVG 44

RESULT 13
Q9X618 PRELIMINARY; PRT; 57 AA.

AC Q9X618;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ALANYL-TRNA SYNTHETASE (FRAGMENT).
GN ALAS.
OS Cytophaga aquatilis.
OC Bacteria; CFb group; Flavobacteriia; Flavobacteriaceae; Flavobacterium.
OX NCBI_TaxID=991;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC29551;
RX MEDLINE:99058143; PubMed:9841678;
RA Gupta R.S.;
RT "Protein phylogenies and signature sequences: A reappraisal of
RT evolutionary relationships among archaeobacteria, eubacteria, and
RT eukaryotes.";
PL Microbiol. Mol. Biol. Rev. 62:1435-1491(1998).

FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC29551;
RA Gupta R.S.; Mukhtar T.; Singh B.;
RX MEDLINE:99058143; PubMed:9841678;
RT "Evolutionary relationships among photosynthetic prokaryotes
RT (Hellobacterium chlorum, Chloroflexus auranticus, Cyanobacteria,
RT Chlorobium tepidum and Proteobacteria): implications regarding the
RT origin of photosynthesis.";
PL Mol. Microbiol. 0:0-0(1999).
DR EMBL: AFI30452; AAD33470.1;
DR InterPro: IPR002318; trna-synt_2c;
DR Pfam: PF01411; trna-synt_2c; 1;
KW Aminoacyl-tRNA synthetase.
FT NON_TER 1
FT NON_TER 57
SQ SEQUENCE 57 AA; 6388 MW; E55B722F7EFBBE16 CRC64;

Query Match 3.5%; Score 7; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 NDLEEVG 24
| | | | | | | |
DB 38 NDLEEVG 44

RESULT 14
O31026 PRELIMINARY; PRT; 98 AA.

AC O31026;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JAN-1998 (TREMBlrel. 05, Last annotation update)
DE HYPOTHETICAL 11.2 KDA PROTEIN.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=866;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-569B,
RA Kaewrakon P.; Manning P.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR PMBL: AP025462; AAB1983.1;
KW Hypothetical protein.
SQ SEQUENCE 98 AA; 11215 MW; 615AC1889B6D4FB4 CRC64;

Query Match 3.5%; Score 7; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 EKLKGLF 85
| | | | | | | |
DB 20 EKLKGLF 26

RESULT 15


```

CC      OF DOUBLE-STRANDED DNA.
CC      -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY
DR      EMBL: AB032581; BAB13319.1; -.
DR      HSSP: P06982; 1AJ6.
DR      InterPro: IPR001241; DNA_topoisomII.
DR      InterPro: IPR003594; HATPase_c.
DR      InterPro: IPR002936; Toprim.
DR      Pfam: PF00204; DNA_topoisomII.1.
DR      Pfam: PF02518; HATPase_c.1.
DR      Pfam: PF01751; Toprim.1.
DR      PRINTS: PR00418; TP12FAMILY.
DR      ProDom: PD000616; DNA_topoisomII.1.
DR      SMART: SM00433; TOP2C.1.
DR      PROSITE: PS00177; TOPOISOMERASE_II.1.
KW      ATP-binding; Isomerase; Topoisomerase.
FT      NON_TER 1
FT      NON_TER 481
SQ      SEQUENCE 481 AA; 53735 MW; 1E4FD8E3E4EEBA30 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SENDPEKC 122
DB 316 SENDPEKC 323
|||||

RESULT 6
QYFAW9 PRELIMINARY; PPT: 481 AA
AC Q9FAW9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DNA GYRASE B SUBUNIT (FRAGMENT)
GN GYRB.
OS Flexibacter saceti.
OC Bacteria; CFB group; Flexibacter group; Flexibacter.
OX NCBI_TaxID=1004;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IFO 15057;
RA Suzuki M., Yamaguchi K.;
RT "Phylogenetic analysis and taxonomic study of marine Cytophaga like bacteria. Proposal of Haerentibaculum gen. nov. with Haerentibaculum maritimum comb. nov. and Haerentibaculum ovolyticus comb. nov., and two new species."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
OF DOUBLE-STRANDED DNA.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
DR      EMBL: AB032582; BAB13320.1; -.
DR      HSSP: P06982; 1AJ6.
DR      InterPro: IPR001241; DNA_topoisomII.
DR      InterPro: IPR003594; HATPase_c.
DR      Pfam: PF00204; DNA_topoisomII.1.
DR      Pfam: PF02518; HATPase_c.1.
DR      Pfam: PF01751; Toprim.1.
DR      PRINTS: PR00418; TP12FAMILY.
DR      ProDom: PD000616; DNA_topoisomII.1.
DR      SMART: SM00433; TOP2C.1.
DR      PROSITE: PS00177; TOPOISOMERASE_II.1.
KW      ATP-binding; Isomerase; Topoisomerase.
FT      NON_TER 1
FT      NON_TER 481
SQ      SEQUENCE 481 AA; 53517 MW; 1F3AB57936BC10F2 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No 7.7;

```

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SENDPEKC 122
DB 316 SENDPEKC 323
|||||

RESULT 7
QYAOI4 PRELIMINARY; PPT: 481 AA.
AC Q9AOI4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DNA GYRASE B SUBUNIT (FRAGMENT).
GN GYRB.
OS Chitinophaga pinensis.
OC Bacteria; CFB group; Flexibacter group; Chitinophaga.
OX NCBI_TaxID=79329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 2588;
PA Suzuki M., Takajima T., Harayama S., Yamamoto S.;
RT "Diversity of marine Cytophaga-like bacteria: Phylogenetic analysis using gyrB sequences and their carotenoids profiles."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
OF DOUBLE-STRANDED DNA.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
DR      EMBL: AB048186; BAB33153.1; -.
DR      HSSP: P06982; 1AJ6.
DR      InterPro: IPR001241; DNA_topoisomII.
DR      InterPro: IPR003594; HATPase_c.
DR      InterPro: IPR002936; Toprim.
DR      Pfam: PF00204; DNA_topoisomII.1.
DR      Pfam: PF02518; HATPase_c.1.
DR      Pfam: PF01751; Toprim.1.
DR      PRINTS: PR00418; TP12FAMILY.
DR      ProDom: PD000616; DNA_topoisomII.1.
DR      SMART: SM00433; TOP2C.1.
DR      PROSITE: PS00177; TOPOISOMERASE_II.1.
KW      ATP-binding; Isomerase; Topoisomerase.
FT      NON_TER 1
FT      NON_TER 481
SQ      SEQUENCE 481 AA; 53681 MW; 584F3B53879927F2 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SENDPEKC 122
DB 316 SENDPEKC 323
|||||

RESULT 8
Q07440 PRELIMINARY; PPT: 488 AA.
AC Q07440;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYANIDE INSENSITIVE TERMINAL OXIDASE.
GN CIOA OR PA3930.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAO1;
RX MEDLINE=97323403; PubMed=9179851;

```

DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DI 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
 DE QUINOL OXIDASE SUBUNIT 1.
 GN MKR4524
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 PE Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=481;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF303099;
 RA Kanoike T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakayuki N., Shimizu S., Sugimoto M.,
 Takemori C., Yamada M., Tabata S.;
 RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.
 RL DNA Res. 7: 641-648 (2000).
 DR EMBL: AP003004; HARS1159.1;
 DR InterPro: IPR002585; Bac_Uba_Cox.
 DR Pfam: PF01654; Bac_Uba_Cox; 1.
 KW Complete proteome.
 SQ SEQUENCE 453 AA; 49427 MW; 102440E6E4BFA0F CRC64;

Query Match 4.0%; Score 8; DB 16; Length 453;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAPFLEA 8
 DB 100 LTAPFLEA 107
 11111111

RESULT 3
 Q9FAX0 PRELIMINARY: PRT: 465 AA.
 ID 086080
 AC 086080
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DI 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 GN QUINOL OXIDASE SUBUNIT 1 JXIA.
 OS Rhodobacter sphaeroides (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 PE Rhodobacter.
 OX NCBI_TaxID=1063;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-2.4.11;
 RA Mowbray N.J., Gak E., Choudhary M., Oh J., Kaplan S.;
 RT Respiratory pathways of Rhodobacter sphaeroides 2.4.11(T);
 RT Identification and characterization of genes encoding quinol
 oxidases.
 RL FEMS Microbiol. Lett. 192:295-310 (2000).
 DR EMBL: AF084042; AAC34673.1;
 DR InterPro: IPR002585; Bac_Uba_Cox.
 DR Pfam: PF01654; Bac_Uba_Cox; 1.
 SQ SEQUENCE 465 AA; 51335 MW; 509F8091E567C910 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 465;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAPFLEA 8
 DB 101 LTAPFLEA 108
 11111111

RESULT 4
 Q9FAX2 PRELIMINARY: PRT: 481 AA.
 ID 09FAX2
 AC 09FAX2
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DI 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 GN DNA GYRASE B SUBUNIT (FRAGMENT).
 OS Flexibacter japonensis.
 OC Bacteria; CF group; Flexibacter group; Flexibacter.
 OX NCBI_TaxID=104662;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-IFO 16041;
 RA Suzuki M., Yamauchi K.;
 RT Phylogenetic analysis and taxonomic study of marine Cytophaga like
 bacteria. Proposal of Haerentibaculum gen. nov. with Haerentibaculum
 maritimum comb. nov. and Haerentibaculum ovolyticus comb. nov., and
 two new species.
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 OF DOUBLE-STRANDED DNA.
 CC -!- SIMILARITY: RELINGS TO THE TYPE II TOPOLISOMERASE FAMILY.
 DR EMBL: AB032579; BAB13317.1;
 DR HSP: P06982; IAS6.
 DR InterPro: IPR001241; DNA_topoisom.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF00204; DNA_TopoII; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PK00418; TP12FAMILY.
 DR ProDom: PD000616; DNA_topoisom; 1.
 DR SMART: SM00433; Top2c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
 KW ATP-binding; Isomerase; Topoisomerase.
 FT NON_TER 1 481
 FT NON_TER 481 481
 SQ SEQUENCE 481 AA; 53556 MW; 8707FBA30H0A70H1 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 481;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 SENDPEKC 122
 DB 316 SENDPEKC 323
 11111111

RESULT 5
 Q9FAX0 PRELIMINARY: PRT: 481 AA.
 ID 09FAX0
 AC 09FAX0
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DI 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 GN DNA GYRASE B SUBUNIT (FRAGMENT).
 OS Flexibacter illinois.
 OC Bacteria; CF group; Flexibacter group; Flexibacter.
 OX NCBI_TaxID=104663;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-IFO 15056;
 RA Suzuki M., Yamauchi K.;
 RT Phylogenetic analysis and taxonomic study of marine Cytophaga like
 bacteria. Proposal of Haerentibaculum gen. nov. with Haerentibaculum
 maritimum comb. nov. and Haerentibaculum ovolyticus comb. nov., and
 two new species.
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:18:30, Search time 96.19 Seconds

(without alignments)
361 493 Million cell updates/sec

Title: US-09-603-665-5_Copy_1867_2067

Perfect score: 201

Sequence: 1 LTAFPLEALDFAHQSHNDL

CTAQSVMAMDSILWKPLNY 201

Scoring table: OLIGO

Gapop 60 0, Gapext 60 0

Searched: 562222 seqs, 17294929 residues

Word size: 6

Total number of hits satisfying chosen parameters: 1169

Minimum BH seq length: 0

Maximum BH seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mbc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	BH	ID	Description
1	100	49.8	349	4	Q96ES5	Q96ES5 homo sapien
2	8	4.0	453	16	Q98DW1	Q98DW1 rhizobium 1
3	8	4.0	465	2	Q8608J	Q8608J rhodobacter
4	8	4.0	481	2	Q9FAX3	Q9FAX3 flexibacter
5	8	4.0	481	2	Q9FAX0	Q9FAX0 flexibacter
6	8	4.0	481	2	Q9FAW0	Q9FAW0 flexibacter
7	8	4.0	481	2	Q9AQL4	Q9AQL4 chitinophag
8	8	4.0	488	16	Q07440	Q07440 pseudomonas
9	8	4.0	493	16	Q92567	Q92567 rhizobium m
10	8	4.0	725	5	Q95X44	Q95X44 caenorhabdi
11	8	4.0	895	5	Q21325	Q21325 caenorhabdi
12	7	3.5	56	2	Q9X617	Q9X617 flexobacter
13	7	3.5	57	2	Q9X618	Q9X618 cytophaga a
14	7	3.5	98	2	Q31026	Q31026 vibrio chol
15	7	3.5	94	11	Q9CNP2	Q9CNP2 mus musculus
16	7	3.5	167	17	Q96Z15	Q96Z15 sulfobus

ALIGNMENTS

RESULT 1

Q96ES5 ID Q96ES5 PRELIMINARY; PRT; 349 AA.
AC Q96ES5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO HYPOPHETICAL PROTEIN FLJ10359.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY, AND ADENOCARCINOMA;
RA Strausberg R.;
PI Submitted (JUL 2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC011983; AHH1983.1; -;
SQ SEQUENCE 349 AA; 39921 MW; 3A359597FF7079EB CRC64;

Query Match

Best Local Similarity 49.8%; Score 100; DB 4; Length 349;

Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTAFPLEALDFAHQSHNDLEFVGKTEHC11DCLVAMVVKLSKVTFRPLFPKLPFWAKTE 60
|||||
DL 72 LTAFPLEALDFAHQSHNDLEFVGKTEHC11DCLVAMVVKLSKVTFRPLFPKLPFWAKTE 131
QY 61 DAPKDRLLTFYNLADICIAEKLKGLTFLFAGHLVKPFPADTL 100
|||||
DL 132 DAPKDRLLTFYNLADICIAEKLKGLTFLFAGHLVKPFPADTL 171

RESULT 2

Q98DW1 ID Q98DW1 PRELIMINARY; PRT; 453 AA.
AC Q98DW1;
DT 01-OCT 2001 (TREMBLrel. 18, Created)

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 HLVKPPA 14
 |||||
 DB 10 HLVKPPA 16

RESULT 15
 US-10-041-018-264
 ; Sequence 264; Application US/10041018
 ; GENERAL INFORMATION:
 ; APPLICANT: Matsuda, Seiichi P.T.
 ; APPLICANT: Hart, Elizabeth A.
 ; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
 ; FILE REFERENCE: P02080US1/10025547
 ; CURRENT APPLICATION NUMBER: US/10/041,018
 ; CURRENT FILING DATE: 2002-01-07
 ; PRIOR APPLICATION NUMBER: US 60/259880
 ; PRIOR FILING DATE: 2001-01-05
 ; NUMBER OF SEQ ID NOS: 413
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO: 264
 ; LENGTH: 1045
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-041-018-264

Query Match 3.6%; Score 7; DB 6; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 HLVKPPA 14
 |||||
 DB 10 HLVKPPA 16

Search completed: July 2, 2002, 16:16:48
 Job time: 825 sec

RESULT 10
US 60 460 049-2667
; Sequence 2667, Application US/60360049
; GENERAL INFORMATION:
; APPLICANT: Cao, Youwei
; APPLICANT: Chen, Xiantong
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-1052052A
; CURRENT APPLICATION NUMBER: US/60/460,049
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2667
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Synchytrium sp.
US 60 460 049-2667

Query Match 4.6%; Score 7; DB 7; Length 415;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 PLATIME 166
|||||

DB 471 PLATIME 477

RESULT 11
US 60 460 049-8006
; Sequence 8006, Application US/60360049
; GENERAL INFORMATION:
; APPLICANT: Cao, Youwei
; APPLICANT: Chen, Xiantong
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-1052052A
; CURRENT APPLICATION NUMBER: US/60/460,049
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8006
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Rhodospirillum rubrum
US 60 460 049-8006

Query Match 4.6%; Score 7; DB 7; Length 852;
Best Local Similarity 100.0%; Pred. No. 1,2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VLALAEK 145
|||||

DB 474 VLALAEK 480

RESULT 12
US 60 460 049-9821
; Sequence 9821, Application US/60360049
; GENERAL INFORMATION:
; APPLICANT: Cao, Youwei
; APPLICANT: Chen, Xiantong
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-1052052A
; CURRENT APPLICATION NUMBER: US/60/460,049
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9821
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Desulfotolbertium halotense
US 60 460 049-9821

Query Match 3.6%; Score 7; DB 7; Length 894;
Best Local Similarity 100.0%; Pred. No. 1,2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VLALAEK 145
|||||

DB 508 VLALAEK 514

RESULT 13
US 10 041 018-199
; Sequence 199, Application US/10041018
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080051/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 199
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US 10 041 018-199

Query Match 3.6%; Score 7; DB 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1,4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 HLKPPA 14
|||||

DB 10 HLKPPA 16

RESULT 14
US 10 041 018-218
; Sequence 218, Application US/10041018
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080051/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 218
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US 10 041 018-218

Query Match 3.6%; Score 7; DB 6; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1,4e+02;

```

; APPLICANT: Ashkenazi, Avi J
; APPLICANT: Baker, Kevin P
; APPLICANT: Bolstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tomas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PLC48 US/09/491,150
; CURRENT APPLICATION NUMBER:
; CURRENT FILING DATE: 2001-11-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 532
; SEQ ID NO 8
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-991-150-8

```

Query Match 3.6%; Score 7; DB 5; Length 367;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LAEKLKE 148
 Db 356 LAEKLKE 362

```

RESULT 7
; Sequence 685, Application US/10137337
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM02BCIN
; CURRENT APPLICATION NUMBER: US/10/137,337
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: 09/58,447
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 812
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 685
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (66)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-137-337-685

```

Query Match 3.6%; Score 7; DB 6; Length 392;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LAEKLKE 148
 Db 111 LAEKLKE 117

```

RESULT 8
US-60-360-039-17151
; Sequence 17151, Application US/603600039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Harry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17151
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-60-360-039-17151

```

Query Match 3.6%; Score 7; DB 7; Length 393;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LAEKLKE 148
 Db 272 LAEKLKE 278

```

RESULT 9
US-10-155-881-27684
; Sequence 27684, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lotfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSPIRATION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 27684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-27684

```

Query Match 3.6%; Score 7; DB 6; Length 409;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 TVLALAE 144
 Db 335 TVLALAE 341

ORGANISM: Caenorhabditis elegans
US-60-460 039 6822

Query Match 3.6% Score 7; DB 7; Length 317;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VLALAEK 145
|||||
DB 244 VLALAEK 249

RESULT 4

PCT-US02-09921-696
; Sequence 696, Application PC/TUS0209921
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DAFEO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Phuong B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GRETZEN, Barry L.
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerald R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: THASON, Olivia
; APPLICANT: YAP, Pierre G.
; APPLICANT: ANSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Susan C.
; APPLICANT: DAM, Lam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Tuy Viet An
; APPLICANT: KEEFEELD, Yael
; APPLICANT: GERSTEIN JR., Edward H.
; APPLICANT: PERALIA, Carolyn H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MAHWAHA, Kakash
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHIRA, Michael
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PF 1242 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/099421
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,290; 60/291,849;
60/291,850; 60/294,436; 60/294,437; 60/294,438
; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
2001-05-17; 2001-06-19; 2001-09-20; 2001-09-20
; NUMBER OF SEQ ID NOS: 1146
; SOFTWARE: PERL Program
; SEQ ID NO 696
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: miscellaneous
; OTHER INFORMATION: Incyte ID No: LG:238456.10.or13-2001MAR40
PCT-US02-09921-696

Query Match 3.6% Score 7; DB 1; Length 352;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LAPKIKE 148
|||||

DB 341 LAEKIKE 347

RESULT 4

US-10-106-698-4568
; Sequence 4568, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005pl
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4568
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4568

Query Match 3.6% Score 7; DB 6; Length 353;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VLALAEK 145
|||||
DB 262 VLALAEK 268

RESULT 5

PCT-US02-09785-698
; Sequence 698, Application PC/TUS0209785
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: US905PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09785
; PRIOR FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 698
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09785-698

Query Match 3.6% Score 7; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TVLGEPL 191
|||||
DB 151 TVLGEPL 157

RESULT 6

US-09-991-150-8
; Sequence 8, Application US/09991150
; GENERAL INFORMATION:

Genware version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:16:47 ; Search time 50.99 seconds
(without alignments)
328.355 Million cell updates/sec

Title: US-09-603-665-5_copy_1950_2144
Perfect score: 195
Sequence: 1 LFTLFGHLVRLPPFADTLXQV CORTIQQLTVLGLPQLSYF 195

Scoring table: Oligo
Gap: 60 0 0 Gap: 60 0 0

Searched: 307805 seqs, 10101595 residues

Word size: 6

Total number of hits satisfying chosen parameters: 376

Minimum DB seq length: 0
Maximum DB seq length: 20000000

Post processing: listing first 45 summaries

- Database:
- 1: Pending Patents, AA, New, *
 - 2: /can2_6/plodata1/paa/pct_NEW_COMB pep.*
 - 3: /can2_6/plodata1/paa/US06_NEW_COMB pep.*
 - 4: /can2_6/plodata1/paa/US07_NEW_COMB pep.*
 - 5: /can2_6/plodata1/paa/US08_NEW_COMB pep.*
 - 6: /can2_6/plodata1/paa/US09_NEW_COMB pep.*
 - 7: /can2_6/plodata1/paa/US10_NEW_COMB pep.*
 - 8: /can2_6/plodata1/paa/US60_NEW_COMB pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	3.6	307	7	US-60-360-039-21343
2	7	3.6	317	7	US-60-360-039-6822
3	7	3.6	352	1	PCT-US02-09921-696
4	7	3.6	353	6	US-10-106-698-4568
5	7	3.6	354	1	PCT-US02-09785-648
6	7	3.6	367	5	US-09-991-150-8
7	7	3.6	392	6	US-10-137-337-685
8	7	3.6	393	7	US-60-360-039-17151
9	7	3.6	400	6	US-10-155-881-27684
10	7	3.6	415	7	US-60-360-039-2667
11	7	3.6	852	7	US-60-360-039-6066
12	7	3.6	894	7	US-60-360-039-9821
13	7	3.6	1045	6	US-10-041-018-194
14	7	3.6	1045	6	US-10-041-018-218
15	7	3.6	1045	6	US-10-041-018-264
16	7	3.6	1045	6	US-10-041-018-309
17	7	3.6	1045	6	US-10-041-018-336
18	7	3.6	1045	7	US-60-360-039-1836
19	7	3.6	1244	7	US-60-360-039-22781
20	7	3.6	1572	5	US-09-935-625-1963
21	7	3.6	1572	5	US-09-935-625-1969
22	7	3.6	1762	5	US-09-935-625-18337
23	7	3.6	1762	5	US-09-935-625-1958
24	7	3.6	1850	5	US-09-935-625-18446
25	7	3.6	1850	5	US-09-935-625-1957
26	7	3.6	2368	5	US-09-935-625-18335
27	7	3.6	2368	5	US-09-935-625-1963

27	7	3.6	2368	5	US-09-935-625-18341	Sequence 18341, A
28	7	3.6	2647	5	US-09-935-625-1962	Sequence 1962, A
29	7	3.6	2647	5	US-09-935-625-18340	Sequence 18340, A
30	7	3.6	2704	5	US-09-935-625-1861	Sequence 1861, A
31	7	3.6	2704	5	US-09-935-625-18339	Sequence 18339, A
32	6	3.1	16	6	US-10-105-299-4193	Sequence 4193, A
33	6	3.1	18	5	US-09-453-834-240	Sequence 240, App
34	6	3.1	18	5	US-09-453-834-244	Sequence 244, App
35	6	3.1	18	5	US-09-453-841-240	Sequence 240, App
36	6	3.1	18	5	US-09-453-841-244	Sequence 244, App
37	6	3.1	18	6	US-10-099-574A-240	Sequence 240, App
38	6	3.1	18	6	US-10-099-574A-244	Sequence 244, App
39	6	3.1	18	6	US-10-147-849-240	Sequence 240, App
40	6	3.1	18	6	US-10-147-849-244	Sequence 244, App
41	6	3.1	18	6	US-10-147-993-240	Sequence 240, App
42	6	3.1	18	6	US-10-147-993-244	Sequence 244, App
43	6	3.1	60	1	US-09-935-625-1963	Sequence 1963, A
44	6	3.1	60	5	US-09-736-968A-58	Sequence 58, App1
45	6	3.1	60	5	US-09-687-837-67	Sequence 67, App1

ALIGNMENTS

RESULT 1

US-60-360-039-21343

; Sequence 21343, Application US/60360039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)A

; CURRENT APPLICATION NUMBER: US/60/460,039

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 21343

; LENGTH: 307

; TYPE: PRT

; ORGANISM: Archaeoglobus fulgidus

US-60-360-039-21343

Query Match 3.6% Score 7; DB 7; Length 307;

Best Local Similarity 100.0%; Pred. No. 46;

Matches 7; Conservative 0, Mismatches 0, Indels 0, Gaps 0;

UY 142 LAEKLKE 148

DB 21 LAEKLKE 27

RESULT 2

US-60-360-039-6822

; Sequence 6822, Application US/60360039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)A

; CURRENT APPLICATION NUMBER: US/60/360,039

; CURRENT FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 6822

; LENGTH: 317

; TYPE: PRT


```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-879-957-79

```

```

Query Match          3.6%, Score 7; DB 22; Length 41;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 33 ENDPEKC 39
   |||||
Db 29 ENDPEKC 35

```

```

RESULT 14
US-09-758-451-174
; Sequence 174, Application US/09758451
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM018
; CURRENT APPLICATION NUMBER: US/09/758,451
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 316
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 174
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-451-174

```

```

Query Match          3.6%, Score 7; DB 21; Length 48;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 41 LLLQFIL 47
   |||||
Db 22 LLLQFIL 28

```

```

RESULT 15
US-09-621-976-4290
; Sequence 4290, Application US/09621976
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4290
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-4290

```

```

Query Match          3.6%, Score 7; DB 20; Length 61;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 147 KENYIVL 153
   |||||
Db 30 KENYIVL 36

```

```

Search completed: July 2, 2002, 16:14:40
Job time: 697 sec

```

1 PRIOR APPLICATION NUMBER: PCT/US01/00667
 2 PRIOR FILING DATE: 2001-01-30
 3 PRIOR APPLICATION NUMBER: PCT/US01/00664
 4 PRIOR FILING DATE: 2001-01-30
 5 PRIOR APPLICATION NUMBER: PCT/US01/00669
 6 PRIOR FILING DATE: 2001-01-30
 7 PRIOR APPLICATION NUMBER: PCT/US01/00665
 8 PRIOR FILING DATE: 2001-01-30
 9 PRIOR APPLICATION NUMBER: PCT/US01/00664
 10 PRIOR FILING DATE: 2001-01-30
 11 PRIOR APPLICATION NUMBER: PCT/US01/00664
 12 PRIOR FILING DATE: 2001-01-30
 13 PRIOR APPLICATION NUMBER: PCT/US01/00662
 14 PRIOR FILING DATE: 2001-01-30
 15 PRIOR APPLICATION NUMBER: PCT/US01/00661
 16 PRIOR FILING DATE: 2001-01-30
 17 PRIOR APPLICATION NUMBER: PCT/US01/00670
 18 PRIOR FILING DATE: 2001-01-30
 19 PRIOR APPLICATION NUMBER: US 60/234,687
 20 PRIOR FILING DATE: 2000-09-21
 21 PRIOR APPLICATION NUMBER: US 09/608,408
 22 PRIOR FILING DATE: 2000-06-30
 23 PRIOR APPLICATION NUMBER: US 09/774,203
 24 PRIOR FILING DATE: 2001-01-29
 25 NUMBER OF SEQ ID NOS: 4117
 26 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 27 SEQ ID NO 4494
 28 LENGTH: 46
 29 TYPE: PRT
 30 ORGANISM: Homo sapiens
 31 FEATURE:
 32 OTHER INFORMATION: MAP TO AC005229.1
 33 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
 34 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9
 35 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.86
 36 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
 37 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
 38 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
 39 US: 09-864 761-44494

Query Match 3.68; Score 7; DB 22; Length 46;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 GREEKQ 49
 DB 4 GREEKQ 10

RESULT 12
 US: 08 417 812A-42
 1 Sequence 42: Application US/08417872A
 2 GENERAL INFORMATION:
 3 APPLICANT: SPARKS, ANDREW H
 4 APPLICANT: HOFFMAN, NOAH
 5 APPLICANT: KAY, BRIAN K.
 6 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN
 7 TITLE OF INVENTION: OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME
 8 NUMBER OF SEQUENCES: 89
 9 CORRESPONDENCE ADDRESS:
 10 ADDRESSEE: Pennie & Edmonds
 11 STREET: 1155 Avenue of the Americas
 12 CITY: New York
 13 STATE: New York
 14 COUNTRY: U.S.
 15 ZIP: 10036
 16 COMPUTER READABLE FORM:
 17 MEDIUM TYPE: Floppy disk
 18 COMPUTER: IBM PC compatible
 19 OPERATING SYSTEM: PC-DOS/MS-DOS
 20 SOFTWARE: Patent In Release #1.0, Version #1.25
 21 CURRENT APPLICATION DATA:
 22 APPLICATION NUMBER: US 08/530,915
 23 FILING DATE: 03-APR-1996
 24 ATTORNEY/AGENT INFORMATION:
 25 NAME: MISTOCK, S. Leslie
 26 REGISTRATION NUMBER: 18,872
 27 REFERENCE/DOCKET NUMBER: 1101-174
 28 TELECOMMUNICATION INFORMATION:
 29 TELEPHONE: (212) 790-9090
 30 TELEFAX: (212) 866-8864/9741
 31 TELEX: 66141 PENNIE
 32 INFORMATION FOR SEQ ID NO: 79:

1 PRIOR APPLICATION NUMBER: US/08417872A
 2 FILING DATE: 07-APR-1995
 3 CLASSIFICATION: 536
 4 ATTORNEY/AGENT INFORMATION:
 5 NAME: MISTOCK, S. Leslie
 6 REGISTRATION NUMBER: 18,872
 7 REFERENCE/DOCKET NUMBER: 1101-193
 8 TELECOMMUNICATION INFORMATION:
 9 TELEPHONE: (212) 790-9090
 10 TELEFAX: 212-866-9741/8664
 11 TELEX: 66141 PENNIE
 12 INFORMATION FOR SEQ ID NO: 32:
 13 SEQUENCE CHARACTERISTICS:
 14 LENGTH: 41 amino acids
 15 TYPE: amino acid
 16 STRANDEDNESS: single
 17 TOPOLOGY: linear
 18 MOLECULE TYPE: protein
 19 HYPOTHEICAL: NO
 20 ANTI-SENSE: NO
 21 US: 08 417-872A-32

Query Match 3.68; Score 7; DB 8; Length 41;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ENDPEK 39
 DB 29 ENDPEK 35

RESULT 13

1 US: 09-879-957-79
 2 Sequence 79: Application US/09879957
 3 GENERAL INFORMATION:
 4 APPLICANT: SPARKS, ANDREW B.
 5 APPLICANT: HOFFMAN, NOAH
 6 APPLICANT: KAY, BRIAN K.
 7 APPLICANT: MCCONNELL, Stephen J.
 8 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 9 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 10 USING SAME
 11 NUMBER OF SEQUENCES: 227
 12 CORRESPONDENCE ADDRESS:
 13 ADDRESSEE: Pennie & Edmonds LLP
 14 STREET: 1155 Avenue of the Americas
 15 CITY: New York
 16 STATE: New York
 17 COUNTRY: USA
 18 ZIP: 10036-2711
 19 COMPUTER READABLE FORM:
 20 MEDIUM TYPE: Floppy disk
 21 COMPUTER: IBM PC compatible
 22 OPERATING SYSTEM: PC-DOS/MS-DOS
 23 SOFTWARE: Patent In Release #1.0, Version #1.30
 24 CURRENT APPLICATION DATA:
 25 APPLICATION NUMBER: US/09,879,957
 26 FILING DATE: 13-Jun-2001
 27 CLASSIFICATION: <unknown>
 28 PRIOR APPLICATION DATA:
 29 APPLICATION NUMBER: US 08/530,915
 30 FILING DATE: 03-APR-1996
 31 ATTORNEY/AGENT INFORMATION:
 32 NAME: MISTOCK, S. Leslie
 33 REGISTRATION NUMBER: 18,872
 34 REFERENCE/DOCKET NUMBER: 1101-174
 35 TELECOMMUNICATION INFORMATION:
 36 TELEPHONE: (212) 790-9090
 37 TELEFAX: (212) 866-8864/9741
 38 TELEX: 66141 PENNIE
 39 INFORMATION FOR SEQ ID NO: 79:

RESULT 8
US-09-823-823-6
; Sequence 6, Application US/09823823
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Satoshi
; APPLICANT: Kasai, Hiroaki
; APPLICANT: Nakamura, Shoko
; APPLICANT: Suzuki, Makoto
; APPLICANT: Hamada, Tohru
; TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS USING G
; FILE REFERENCE: 12817-004001
; CURRENT APPLICATION NUMBER: US/09/823-823
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/208,688
; PRIOR FILING DATE: 1998-12-10
; PRIOR APPLICATION NUMBER: JP 97/343316
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 6
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Chitinophaga pinensis
US-09-823-823-6

Query Match 4.1%, Score 8, DB 22, Length 220,
Best Local Similarity 100.0%, Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SENDPEKC 39
| | | | | | | |
DB 130 SENDPEKC 137

RESULT 9
US-09-823-829-6
; Sequence 6, Application US/09823829
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Satoshi
; APPLICANT: Nakamura, Shoko
; APPLICANT: Suzuki, Makoto
; APPLICANT: Kasai, Hiroaki
; APPLICANT: Hamada, Tohru
; TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS
; FILE REFERENCE: 12817-004001
; CURRENT APPLICATION NUMBER: US/09/823-829
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/208,688
; PRIOR FILING DATE: 1998-12-10
; PRIOR APPLICATION NUMBER: JP 97/343316
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 6
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Chitinophaga pinensis
US-09-823-829-6

Query Match 4.1%, Score 8, DB 22, Length 220;
Best Local Similarity 100.0%, Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SENDPEKC 39
| | | | | | | |
DB 130 SENDPEKC 137

RESULT 10

PCT-US01-00663-34446
; Sequence 34446, Application PC/TUS0100663
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: PB 0004 WO 7
; CURRENT APPLICATION NUMBER: PCT/US01/00663
; CURRENT FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 64 February 2000 (04 02 00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26 05 00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03 08 00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03 10 00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27 09 00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21 09 00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30 06 00)
; NUMBER OF SEQ ID NOS: 38837
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 34446
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005229.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 0.96
PCT-US01-00663-34446

Query Match 3.6%, Score 7, DB 1, Length 36;
Best Local Similarity 100.0%, Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 G3EEKEQ 89
| | | | | | | |
DB 4 G3EEKEQ 10

RESULT 11
US-09-864-761-44493
; Sequence 44493, Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30

```

? PRIOR FILING DATE: 2001-12-22
? PRIOR APPLICATION NUMBER: US 09/491,404
? PRIOR FILING DATE: 2000-01-25
? PRIOR APPLICATION NUMBER: PCT/US01/02624
? PRIOR FILING DATE: 2001-01-25
? PRIOR APPLICATION NUMBER: US 09/496,914
? PRIOR FILING DATE: 2000-02-03
? PRIOR APPLICATION NUMBER: US 09/560,875
? PRIOR FILING DATE: 2000-04-27
? PRIOR APPLICATION NUMBER: PCT/US01/03800
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 09/515,126
? PRIOR FILING DATE: 2000-02-28
? PRIOR APPLICATION NUMBER: US 09/577,409
? PRIOR FILING DATE: 2000-05-18
? PRIOR APPLICATION NUMBER: PCT/US01/04367
? PRIOR FILING DATE: 2001-02-26
? PRIOR APPLICATION NUMBER: US 09/519,705
? PRIOR FILING DATE: 2000-03-07
? PRIOR APPLICATION NUMBER: US 09/574,454
? PRIOR FILING DATE: 2000-05-19
? PRIOR APPLICATION NUMBER: PCT/US01/04941
? PRIOR FILING DATE: 2001-03-05
? PRIOR APPLICATION NUMBER: US 09/540,217
? PRIOR FILING DATE: 2000-03-31
? PRIOR APPLICATION NUMBER: US 09/649,167
? PRIOR FILING DATE: 2000-08-24
? PRIOR APPLICATION NUMBER: PCT/US01/08641
? PRIOR FILING DATE: 2001-03-30
? PRIOR APPLICATION NUMBER: US 09/552,929
? PRIOR FILING DATE: 2000-04-18
? PRIOR APPLICATION NUMBER: US 09/770,160
? PRIOR FILING DATE: 2001-01-26
? PRIOR APPLICATION NUMBER: PCT/US01/08656
? PRIOR FILING DATE: 2001-04-18
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: P1_Flucenes Version 6.0
? SEQ ID NO 167
? LENGTH: 1149
? TYPE: PRT
? ORGANISM: Homo sapiens
US_60_339_453_167

```

```

Query Match          65.1%; Score 127; DB 26; Length 1149;
Best Local Similarity 100.0%; Pred. No. 4.6e-119;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALMPPLVDQLENPLGGKFKFQPVTKHITPCIALQFSVAMADDSIWKPLNYQILLKTPDSS 128
DB 1023 ALMPPLVDQLENPLGGKFKFQPVTKHITPCIALQFSVAMADDSIWKPLNYQILLKTPDSS 1082

QY 129 PKVRFALITVIALAEKLNKXNYIVLLPESIPFLAEIMEDIEVEVEHQCTIQQLLETIVLG 188
DB 1083 PKVRFALITVIALAEKLNKXNYIVLLPESIPFLAEIMEDIEVEVEHQCTIQQLLETIVLG 1142

QY 189 EPLQSYF 195
DB 1143 EPLQSYF 1149

```

```

RESULT 7
US_60_141_424_4
? Sequence 4; Application US/60141424
? GENERAL INFORMATION:
? APPLICANT: Batt, Caroline
? APPLICANT: Bouquelerot, Lydie
? TITLE OF INVENTION: A NOVEL BAP2H GENE AND PROTEIN
? FILE REFERENCE: GENSET 0539F
? CURRENT APPLICATION NUMBER: US 09/234,325
? PRIOR FILING DATE: 1999-06-25
? NUMBER OF SEQ ID NOS: 5

```

```

? SOFTWARE: Patent.pm
? SEQ ID NO 3
? LENGTH: 2144
? TYPE: PRT
? ORGANISM: homo sapiens
? NAME/KEY: TRANSMEM
? LOCATION: 91..111
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 532..532
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 700..720
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 861..881
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 911..931
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 937..957
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 1297..1317
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 1410..1430
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 1725..1745
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: TRANSMEM
? LOCATION: 1987..2007
? OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
? NAME/KEY: SITE
? LOCATION: 1750..1771
? OTHER INFORMATION: potential leucine zipper pattern, BLA Program
US_60_141-323-3

```

```

Query Match          65.1%; Score 127; DB 26; Length 2144;
Best Local Similarity 100.0%; Pred. No. 7.8e-119;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALMPPLVDQLENPLGGKFKFQPVTKHITPCIALQFSVAMADDSIWKPLNYQILLKTPDSS 128
DB 2018 ALMPPLVDQLENPLGGKFKFQPVTKHITPCIALQFSVAMADDSIWKPLNYQILLKTPDSS 2077

QY 129 PKVRFALITVIALAEKLNKXNYIVLLPESIPFLAEIMEDIEVEVEHQCTIQQLLETIVLG 188
DB 2078 PKVRFALITVIALAEKLNKXNYIVLLPESIPFLAEIMEDIEVEVEHQCTIQQLLETIVLG 2147

QY 189 EPLQSYF 195
DB 2138 EPLQSYF 2144

```

QY 129 PKVRFALITVLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQQLTETVLG 188
DB 449 PKVRFALITVLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQQLTETVLG 508
QY 189 EPLQSYF 195
DB 509 EPLQSYF 515

RESULT 4

US-60-042-611-52
; Sequence 52, Application US/60042611
; GENERAL INFORMATION:
; APPLICANT: Howcock, Anne M.
; APPLICANT: Baer, Richard
; TITLE OF INVENTION: Compositions and Methods Comprising
; TITLE OF INVENTION: BARD1 and Other BRCA1 Binding Proteins
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESS: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/042,611
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: P-41,071
; REFERENCE/DOCKET NUMBER: UTSD:499PZ2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-60-042-611-52

Query Match 65.1%, Score 127; DB 26; Length 515;
Best Local Similarity 100.0%, Pred. No. 2,2e-119;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALMMPLVDQLENRLGSGEEKFQERVTKHILPCIAQFSVAMADSLWKFLNYOILLKTRDSS 128
DB 389 ALMMPLVDQLENRLGSGEEKFQERVTKHILPCIAQFSVAMADSLWKFLNYOILLKTRDSS 448
QY 129 PKVRFALITVLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQQLTETVLG 188
DB 449 PKVRFALITVLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQQLTETVLG 508
QY 189 EPLQSYF 195
DB 509 EPLQSYF 515

RESULT 5

US-60-042-985-52
; Sequence 52, Application US/60042985
; GENERAL INFORMATION:
; APPLICANT: Howcock, Anne M.
; APPLICANT: Baer, Richard

; TITLE OF INVENTION: Compositions and Methods Comprising
; TITLE OF INVENTION: BARD1 and Other BRCA1 Binding Proteins
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESS: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/042,985
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: P-41,071
; REFERENCE/DOCKET NUMBER: UTSD:499PZ3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-60-042-985-52

Query Match 65.1%, Score 127; DB 26; Length 515;
Best Local Similarity 100.0%, Pred. No. 2,2e-119;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALMMPLVDQLENRLGSGEEKFQERVTKHILPCIAQFSVAMADSLWKFLNYOILLKTRDSS 128
DB 389 ALMMPLVDQLENRLGSGEEKFQERVTKHILPCIAQFSVAMADSLWKFLNYOILLKTRDSS 448
QY 129 PKVRFALITVLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQQLTETVLG 188
DB 449 PKVRFALITVLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQQLTETVLG 508
QY 189 EPLQSYF 195
DB 509 EPLQSYF 515

RESULT 6

US-60-339-453-167
; Sequence 167, Application US/60339453
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Chosh, Malabika
; APPLICANT: Weng, Gezhi
; APPLICANT: Hoyle, Bryan J
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 812
; CURRENT APPLICATION NUMBER: US/60/339,453
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000 04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017

LOCATION: 1967
OTHER INFORMATION: Xaa-Asp or Asn
NAME/KEY: VARIANT
LOCATION: 2017
OTHER INFORMATION: Xaa-Gly or Glu
US-09-603-665-5

Query Match 99.0% Score 193; DB 20; Length 2144;
Best Local Similarity 100.0%; Pred. No. 1e-185;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTLFAHILVKPPADTLKXVNI SKTDEAF FUSENPEKCKLLIQFLNLYKIFLEDTQH 609
DB 1950 LFTLFAHILVKPPADTLKXVNI SKTDEAF FUSENPEKCKLLIQFLNLYKIFLEDTQH 2009
QY 61 FISKEAXALMPLVLTQLENPLGSGEEKFQFFVTKHILIPICIAQFSVAMALDLSLWKPLNYQI 129
DB 2010 FISKEAXALMPLVLTQLENPLGSGEEKFQFFVTKHILIPICIAQFSVAMALDLSLWKPLNYQI 2069
QY 121 LKTRDSSPKVRFALITVLALAEKLEKENVIVLLPESIPFTLAELMEDECEVEEHQCKTI 180
DB 2070 LKTRDSSPKVRFALITVLALAEKLEKENVIVLLPESIPFTLAELMEDECEVEEHQCKTI 2129
QY 181 QLETVLGEPLQSYF 195
DB 2140 QLETVLGEPLQSYF 2144

RESULT 2
US 60-176-880-4
Sequence 4; Application US/60176880
GENERAL INFORMATION:
APPLICANT: Barry, Caroline
APPLICANT: Bouqueleret, Lydie
APPLICANT: Choumakov, Ilya
TITLE OF INVENTION: A NOVEL HAP28 GENE AND PROTEIN
FILE REFERENCE: 67 US2,PRO
CURRENT APPLICATION NUMBER: US/60-176-880
CURRENT FILING DATE: 2000-01-18
EARLIER APPLICATION NUMBER: US 60/141,323
EARLIER FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patent.ram
SEQ ID NO 4
LENGTH: 2144
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: 1694
OTHER INFORMATION: Xaa-Ser or Asn
FEATURE:
NAME/KEY: VARIANT
LOCATION: 1854
OTHER INFORMATION: Xaa-Ala or Val
FEATURE:
NAME/KEY: VARIANT
LOCATION: 1967
OTHER INFORMATION: Xaa-Asp or Asn
FEATURE:
NAME/KEY: VARIANT
LOCATION: 2017
OTHER INFORMATION: Xaa-Gly or Glu
US 60-176-880-4

Query Match 99.0% Score 193; DB 26; Length 2144;
Best Local Similarity 100.0%; Pred. No. 1e-185;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTLFAHILVKPPADTLKXVNI SKTDEAF FUSENPEKCKLLIQFLNLYKIFLEDTQH 609
DB 1950 LFTLFAHILVKPPADTLKXVNI SKTDEAF FUSENPEKCKLLIQFLNLYKIFLEDTQH 2009

DB 1950 LFTLFAHILVKPPADTLKXVNI SKTDEAF FUSENPEKCKLLIQFLNLYKIFLEDTQH 2009
QY 61 FISKEAXALMPLVLTQLENPLGSGEEKFQFFVTKHILIPICIAQFSVAMALDLSLWKPLNYQI 120
DB 2010 FISKEAXALMPLVLTQLENPLGSGEEKFQFFVTKHILIPICIAQFSVAMALDLSLWKPLNYQI 2069
QY 121 LKTRDSSPKVRFALITVLALAEKLEKENVIVLLPESIPFTLAELMEDECEVEEHQCKTI 180
DB 2070 LKTRDSSPKVRFALITVLALAEKLEKENVIVLLPESIPFTLAELMEDECEVEEHQCKTI 2129
QY 181 QLETVLGEPLQSYF 195
DB 2140 QLETVLGEPLQSYF 2144

RESULT 3
US-08-936-487-52
Sequence 52; Application US/08936487
GENERAL INFORMATION:
APPLICANT: Howcock, Anne M.
APPLICANT: Baer, Richard
TITLE OF INVENTION: Compositions and Methods Comprising
TITLE OF INVENTION: BARD1 and Other BARD1 Binding Proteins
NUMBER OF SEQUENCES: 130
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P O Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1 0, Version #1 30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/CB/936,487
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/025,296
FILING DATE: 20-SEP-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/042,611
FILING DATE: 03-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/042,985
FILING DATE: 04-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W.
REGISTRATION NUMBER: P-41,071
REFERENCE/POWER NUMBER: UTS-499
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 515 amino acids
TYPE: amino acid
STRANDNESS:
TOPOLOGY: linear
US-08-936-487-52

Query Match 65.1% Score 127; DB 13; Length 515;
Best Local Similarity 100.0%; Pred. No. 2.2e-119;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALMPLVLTQLENPLGSGEEKFQFFVTKHILIPICIAQFSVAMALDLSLWKPLNYQILKTRDSS 128
DB 389 ALMPLVLTQLENPLGSGEEKFQFFVTKHILIPICIAQFSVAMALDLSLWKPLNYQILKTRDSS 448

genCore version 4.5
Copyright (C) 1993 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:14:39 : Search time 50% 3 seconds

(without alignments)
135.833 Million cell updates/sec

Title: US-09-603-665-5 COPY_1950_2144

Perfect score: 195
Sequence: 1 LFTLFAHLVLPFAFTXQV

Scoring table: Oligo Gapop 6.0 0.0 Gapext 6.0 0.0

Searched: 350226 seqs, 351980561 residues

Word size: 6

Total number of hits satisfying chosen parameters: 2299

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Pending Patents AA.Main:

- 1: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 2: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 3: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 4: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 5: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 6: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 7: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 8: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 9: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 10: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 11: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 12: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 13: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 14: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 15: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 16: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 17: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 18: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 19: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 20: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 21: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 22: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 23: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 24: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 25: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*
- 26: /cqn2_6/ptodata/2/paa/us086_COMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	193	99.0	2144	20	US-09-603-665-5	Sequence 5, Appli
2	164	99.0	2144	26	US-60-176-880-4	Sequence 4, Appli
3	127	65.1	515	13	US-08-936-487-52	Sequence 52, Appl
4	127	65.1	515	26	US-60-042-611-52	Sequence 52, Appl
5	127	65.1	515	26	US-60-042-985-52	Sequence 52, Appl
6	127	65.1	1149	26	US-60-339-453-167	Sequence 167, App
7	127	65.1	2144	26	US-60-141-323-3	Sequence 3, Appli

8	4.1	220	22	US-09-823-823-6	Sequence 6, Appli
9	4.1	220	22	US-09-823-823-6	Sequence 6, Appli
10	7	3.6	46	US-09-864-761-44493	Sequence 44493, A
11	7	3.6	36	US-09-864-761-44493	Sequence 44493, A
12	7	3.6	41	US-08-417-872A-32	Sequence 32, Appl
13	7	3.6	41	US-09-879-957-79	Sequence 79, Appl
14	7	3.6	48	US-09-758-451-174	Sequence 174, App
15	7	3.6	41	US-09-758-451-174	Sequence 174, App
16	7	3.6	61	US-60-147-499-4290	Sequence 4290, Ap
17	7	3.6	89	US-60-196-718-1436	Sequence 7436, Ap
18	7	3.6	93	US-60-196-718-1436	Sequence 7436, Ap
19	7	3.6	97	US-60-162-247-3315	Sequence 3315, Ap
20	7	3.6	97	US-60-162-247-3315	Sequence 3315, Ap
21	7	3.6	98	US-09-751-181-38	Sequence 38, Appl
22	7	3.6	110	US-09-864-761-37303	Sequence 37303, A
23	7	3.6	110	US-09-864-761-37303	Sequence 37303, A
24	7	3.6	111	US-09-107-532-6705	Sequence 6705, Ap
25	7	3.6	111	US-09-107-532-6705	Sequence 6705, Ap
26	7	3.6	111	US-09-107-532A-6705	Sequence 6705, Ap
27	7	3.6	113	US-60-170-373-2898	Sequence 2898, Ap
28	7	3.6	117	US-09-760-479-748	Sequence 748, App
29	7	3.6	124	US-09-760-479-748	Sequence 748, App
30	7	3.6	124	US-09-760-479-748	Sequence 748, App
31	7	3.6	124	US-09-760-479-748	Sequence 748, App
32	7	3.6	134	US-08-625-811-1287	Sequence 1287, Ap
33	7	3.6	134	US-08-625-811-1287	Sequence 1287, Ap
34	7	3.6	140	US-08-993-002A-5497	Sequence 5497, Ap
35	7	3.6	140	US-60-181-418-145	Sequence 145, App
36	7	3.6	148	US-09-897-516-6127	Sequence 6127, Ap
37	7	3.6	148	US-09-897-516-6127	Sequence 6127, Ap
38	7	3.6	150	US-09-732-210-1095	Sequence 1095, Ap
39	7	3.6	150	US-09-732-210-1095	Sequence 1095, Ap
40	7	3.6	164	US-60-140-956-2146	Sequence 2146, Ap
41	7	3.6	178	US-09-339-972-57	Sequence 57, Appl
42	7	3.6	188	US-09-675-784A-13298	Sequence 13298, A
43	7	3.6	189	US-09-708-427-85493	Sequence 85493, A
44	7	3.6	200	US-09-708-427-85493	Sequence 85493, A
45	7	3.6	204	US-09-708-427-85493	Sequence 85493, A
46	7	3.6	211	US-60-142-896-1282	Sequence 1282, Ap

ALIGNMENTS

RESULT 1
US-09-603-665-5
Sequence 5, Application US/09603665
GENERAL INFORMATION:
APPLICANT: Barry, Caroline
AFFILIANT: Bouquellere, Lydie
ATTORNEY: Chumakov, Ilya
TITLE OF INVENTION: A NOVEL BAI-28 GENE AND PROTEIN
FILE REFERENCE: GENSEL US3AUS
CURRENT APPLICATION NUMBER: US/09/603,665
CURRENT FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/141,323
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 60/176,880
PRIOR FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 63
SOFTWARE: Patent.pm
SEQ ID NO 5
LENGTH: 2144
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: 1674
OTHER INFORMATION: Xaa-Ser or Asn
NAME/KEY: VARIANT
LOCATION: 1854
OTHER INFORMATION: Xaa-Ala or Val
NAME/KEY: VARIANT

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schenckebach C., Seva T., Shibata Y., Storch K.-E.,
 Suzuki H., Toyooka K., Wada K.H., Weitz C., Whittaker C., Wilming L.,
 Wyshak Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 Hayashizaki Y.;
 RI *Functional annotation of a full length mouse cDNA collection.*;
 RL Nature 409:685-690(2001).
 TR EMBL: AK011157; BAB27437.1; -.
 TR BSSD: P14550; 2ALR.
 DR MCD: MCL1929955; AKT1A1.
 DR InterPro: IPR001495; Aldo_kof_red.
 DR Pfam: PF00248; aldokof_red: 1.
 DR PRINTS: PR00069; ALDKETRIASE.
 TR PROSITE: PS00062; ALD-KETO_PEROXIDASE_2: 1.
 FI NON-REF 1
 SU SEQUENCE 252 AA: 24869 MW: 8983305A4547108A CR664;

Query Match 3.68; Score 7; DB 11; Length 252;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

<Y 149 VLALAEK 145
 ID 161 VLALAEK 167

Search completed: July 2, 2002, 16:18:33
 Job time: 865 sec

```

Query Match          3.6%; Score 7; DB 11; Length 214;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VLAAEK 145
DB 123 VLAAEK 129
      |||||
      |||||

RESULT 12
Q70473          PRELIMINARY; PRT; 228 AA.
AC Q70473;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ALDEHYDE REDUCTASE (FRAGMENT).
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Hyndman D.J., Flynn T.G.;
RT "Partial sequence of Chinese hamster aldehyde reductase.",
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF060820; AAC15760.1;
DR HSSP: P14550; 2ALR.
DR InterPro: IPR001395; Ald_ket_red
DR Pfam: PF00248; aldo_ket_red; 1.
DR PRINTS: PK00669; ALDKETRDTASE.
DR PROSITE: PS00798; ALDO_KETO_REDUCTASE_1; 1.
DR PROSITE: PS00062; ALDO_KETO_REDUCTASE_2; 1.
FT NON_TER 1
FT NON_TER 228
FT SEQUENCE 228 AA; 25450 MW; 90456255PIR010000AA; rKQ664;

Query Match          3.6%; Score 7; DB 11; Length 228;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VLAAEK 145
DB 219 VLAAEK 225
      |||||
      |||||

RESULT 13
Q90WT1          PRELIMINARY; PRT; 232 AA.
AC Q90WT1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CRYSTALLIN B1 PROTEIN.
GN CRYSTALLIN B1.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen J.-Y.;
RL Thesis (2001), Department of Institute of Zoology, Academia Sinica, Taipei, Taiwan.
DR EMBL: AJ317957; CAC64899.1;
SQ SEQUENCE 232 AA; 26781 MW; 517B0B1A29AA058R rKQ664;

Query Match          3.6%; Score 7; DB 13; Length 232;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VLAAEK 145
DB 219 VLAAEK 225
      |||||
      |||||

Query Match          3.6%; Score 7; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 DSSPKVR 132
DB 11 DSSPKVR 17
      |||||
      |||||

RESULT 15
Q9CT53          PRELIMINARY; PRT; 252 AA.
AC Q9CT53;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 2610201A1APIK PROTEIN (FRAGMENT).
GN AKP1A1 OR 2610201A1APIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=EMBRYO;
BX MEDLIN=21095660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Akizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King H., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli P., Suzuki K., Tonita M., Wagner I., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bojelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons F., Marchionni L., Mashima J., Manzarelli J., Mombaerts P.,

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE:R9124097; PubMed-2644218;
RA Jacobson M.R., Briggs K.E., Bennett L.T., Sotterquist R.A.,
RA Wilson M.S., Cash V.L., Beynon J., Newton W.E., Dean D.R.;
PI *Physical and genetic map of the major rif gene cluster from
RT Azolobacter vinelandii*.
RL J. Bacteriol. 171:1017-1027(1989).
DR EMBL: M20568; AAA64721.1;
SQ SEQUENCE 72 AA; 8604 MW; 60CA2DBE3045PAA CRC64;

Query Match 3.6%; Score 7; DB 2; Length 72;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 ALAEKIK 147
DB 14 ALAEKIK 24
IIIIII

RESULT 4
Q9LZV5 PRELIMINARY; PRT; 165 AA.
AC Q9LZV5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHELICAL 22.5 KDA PROTEIN.
GN L20135_180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplanet; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Cuscuta; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN 111
SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Lasseville P.,
RA de Clerck R., de Kuyser A., Ney P., Ronze P., Van Den Daele H.,
RA Villarroel R., Gielens J., Van Montagu M., Hancock T., Mewes H.W.,
RA Kudd S., Tempe K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases
RN 121
SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases
DR EMBL: AL16235; CAB82561.1;
KW Hypothetical protein.
SQ SEQUENCE 165 AA; 19002 MW; 2140B7CBAA98BA2B CRC64;

Query Match 3.6%; Score 7; DB 10; Length 165;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AEKIKEN 149
DB 30 AEKIKEN 36
IIIIII

RESULT 10
Q90G19 PRELIMINARY; PRT; 205 AA.
AC Q90G19;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHELICAL 22.5 KDA PROTEIN (FRAGMENT).
GN HKZF58661624.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 111

```

```

RP SEQUENCE FROM N.A.
RC Tissue-hoer A., Lauber J., Mewes H.W., Gassonhuber J., Wiemann S.;
RI Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC -1- SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AL110160; CAB53665.1;
DR IPI-IPRO: IPR001690; WD40
DR Pfam: PF00400; WD40; 2.
DR SMART: SM00320; WD40; 2.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS00082; WD_REPEATS_2; 1.
DR PROSITE: PS00244; WD_REPEATS_PESION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT NON_TER 1
SQ SEQUENCE 205 AA; 22537 MW; 972H1FF96E460B10 CRC64;

Query Match 3.6%; Score 7; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 185 TVLGEPL 191
DB 137 TVLGEPL 143
IIIIII

RESULT 11
Q9D0P3 PRELIMINARY; PRT; 214 AA.
AC Q9D0P3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 2610201A18K PROTEIN.
GN AKR1A1 OR 2610201A18K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 111
SEQUENCE FROM N.A.
RC STRAIN-G57BL/6J; TISSUE-EM3RYO;
RX MEDLINE:21085660; PubMed-11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gajjovori T., Hono H., Kasukawa I., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Koehliwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Haidarelli R., Harsh G.,
RA Hlake J., Hoffelli D., Hojunga N., Carninci P., de Hualdo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Holmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring H., Ringwald M., Rodriguez L., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya I., Shibata I., Storch K.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz G., Whitaker G., Wilming L.,
RA Wynshaw-Boris A., Yoshida Y., Yasuda Y., Kawaji H., Kohzuki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RN 111
Nature 409:685-690(2001).
DR EMBL: AK011221; HA827477.1;
DR HSSP: PL4550; 2AIP.
DR MGD: MGI:1929455; Akrlal.
DR InterPro: IPR001305; Aldo_ket_red.
DR Pfam: PF00248; aldo_ket_red; 1.
DR PRINTS: PF00069; ALDRETDIASE.
DR PROSITE: PS00062; ALDO_KETO_PENNYTAG; 2; 1.
SQ SEQUENCE 214 AA; 24419 MW; 62AA49AC146C8E5 CRC64;

```

maritimum comb. nov. and Haerentibaculum ovoliticus comb. nov., and two new species.*
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 DR EMBL: AH032581; HARI3319.1; -.
 DR HSSP: P06982; 1AJ6.
 DR InterPro: IPR001241; DNA_topoisomII.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF00204; DNA_topoisomII; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR ProDom: PD000616; DNA_topoisomII; 1.
 DR SMART: SM00433; TOP2c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 KW ATP-binding; Isomerase; Topoisomerase.
 FT NON_TER 1 1
 FT NON_TER 481 481
 SQ SEQUENCE 481 AA; 53735 MW; 1E4FD8E3F4EEBA30 CRC64;

Query Match 4.1%, Score 8; DB 2, Length 481;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SENDPEKC 39
 |||||
 Db 316 SENDPEKC 323

RESULT 6
 Q9FAW9 ID Q9FAW9 PRELIMINARY; PRT; 481 AA.
 AC Q9FAW9
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE DNA GYRASE B SUBUNIT (FRAGMENT).
 GN GYR.
 OS Flexibacter sancti.
 OC Bacteria; CFH group; Flexibacter group; Flexibacter.
 OX NCBI_TaxID=1004;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IP0 15057.
 RA Suzuki M., Yamaguchi K.;
 RT "Phylogenetic analysis and taxonomic study of marine Cytophaga like bacteria. Proposal of Haerentibaculum gen. nov. with Haerentibaculum maritimum comb. nov. and Haerentibaculum ovoliticus comb. nov., and two new species.*"
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 DR EMBL: AB032582; BAB13320.1; -.
 DR HSSP: P06982; 1AJ6.
 DR InterPro: IPR001241; DNA_topoisomII.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF00204; DNA_topoisomII; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR ProDom: PD000616; DNA_topoisomII; 1.
 DR SMART: SM00433; TOP2c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 KW ATP-binding; Isomerase; Topoisomerase.
 FT NON_TER 1 1
 FT NON_TER 481 481
 SQ SEQUENCE 481 AA; 53517 MW; 1F3AB57936HC10F2 CRC64;

Query Match 4.1%, Score 8; DB 2; Length 481;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SENDPEKC 39
 |||||
 Db 316 SENDPEKC 323

RESULT 7
 Q9AQ14 ID Q9AQ14 PRELIMINARY; PRT; 481 AA.
 AC Q9AQ14
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE DNA GYRASE B SUBUNIT (FRAGMENT).
 GN GYR.
 OS Chitinophaga pinensis.
 OC Bacteria; CFH group; Flexibacter group; Chitinophaga.
 OX NCBI_TaxID=79329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 2588;
 RA Suzuki M., Takadera T., Harayama S., Yamamoto S.;
 RT "Diversity of marine Cytophaga-like bacteria: Phylogenetic analysis using gyrB sequences and their carotenoids profiles.*"
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 DR EMBL: AB048186; BAB33153.1; -.
 DR HSSP: P06982; 1AJ6.
 DR InterPro: IPR001241; DNA_topoisomII.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF00204; DNA_topoisomII; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR ProDom: PD000616; DNA_topoisomII; 1.
 DR SMART: SM00433; TOP2c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 KW ATP-binding; Isomerase; Topoisomerase.
 FT NON_TER 1 1
 FT NON_TER 481 481
 SQ SEQUENCE 481 AA; 53681 MW; E8452458879927F2 CRC64;

Query Match 4.1%, Score 8; DB 2; Length 481;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SENDPEKC 39
 |||||
 Db 316 SENDPEKC 323

RESULT 8
 Q44537 ID Q44537 PRELIMINARY; PRT; 72 AA.
 AC Q44537
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE ORF 11.
 OS Azotobacter vinelandii.
 OC Bacteria; Proteobacteria; gamma subdivision, Pseudomonadaceae;
 OX Azotobacter.
 OX NCBI_TaxID=354;
 RN [1]

Q9K6G3
ID Q9K6G3 PRELIMINARY: PRI: 188 AA.
AC Q9K6G3
DT 01-OCT-2000 (TREMREL: 15, Created)
DI 01-OCT-2000 (TREMREL: 15, Last sequence update)
DE 01-OCT-2000 (TREMREL: 15, Last annotation update)
DE BH466 PROTEIN.
GN BH466
OS Bacteria: Firmicutes: Bacillus/Clostridium group: Bacillus.
OC Bacillus/Staphylococcus group: Bacillus.
OX NCBI_TaxID=86663;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IFO 9253;
RA MEDLINE=20512582; PubMed 11058132.
KA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
KA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,
KA Horikoshi K.;
KW *Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis *;
RL Nucleic Acids Res. 28:4317-4331(2000).
DE EMBL: AP001519; BAB07485.1;
KW Complete proteome.
SQ SEQUENCE 188 AA: 26473 MW: 45208644BD45D47 CRC64;

Query Match 4.1%; Score 8; DB 16; Length 188;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 181 QLELVIG 188
IIIIIIII
6 QLELVIG 14

RESULT 3
Q9K6G3
ID Q9K6G3 PRELIMINARY: PRI: 253 AA.
AC Q9K6G3
DT 01-MAR-2001 (TREMREL: 16, Created)
DI 01-MAR-2001 (TREMREL: 16, Last sequence update)
DE 01-DEC-2001 (TREMREL: 19, Last annotation update)
DE MHC CLASS II ALPHA CHAIN.
GN HCTIALPHA.
OS Calman crocodilus (Speckled calman) (Calman sclerops).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Archosauria: Crocodylidae: Alligatorinae: Calman.
OX NCBI_TaxID=8499;
RN [1]
RP SEQUENCE FROM N.A.
RC Voldby J., Vilved L., Dup M., Gronlund J., Holmskov U., Teisner B.,
RA Salomonsen J., Brunsdard K., Skjold K.;
RA *Cloning, sequence and genomic structure of MHC class II antigens from
RI the speckled calman, Calman crocodilus.*;
RL Submitted (MAR 2000) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF256650; AAF09282.1;
DE HSSP: P01904; 2SEB
DR InterPro: IPR003597; 14_c1.
DR InterPro: IPR004006; 14_MHC.
DR InterPro: IPR001003; MHC_II_alpha.
DR Pfam: PF00047; 14_1
DR Pfam: PF00994; MHC_II_alpha; 1.
DR SMART: SM00407; IGL1_1.
DR PROSITE: PS00290; IG_MHC; UNKN0WN_1.
SQ SEQUENCE 253 AA: 28550 MW: D95EA0D9060E4325 CRC64;

Query Match 4.1%; Score 8; DB 7; Length 253;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 145 ALITVIAL 142

DB 13 ALITVIAL 20
IIIIIIII

RESULT 4
Q9FAX2
ID Q9FAX2 PRELIMINARY: PRI: 481 AA.
AC Q9FAX2
DT 01-MAR-2001 (TREMREL: 16, Created)
DI 01-MAR-2001 (TREMREL: 16, Last sequence update)
DE 01-DEC-2001 (TREMREL: 19, Last annotation update)
DE DNA GYRASE B SUBUNIT (FRAGMENT).
GN GYRB.
OS Flexibacter japonensis.
OC Bacteria: CFB group: Flexibacter group: Flexibacter.
OX NCBI_TaxID=104662;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IFO 16041;
RA Suzuki M., Yamaguchi K.;
PT *Phylogenetic analysis and taxonomic study of marine Cytophaga like
RI bacteria. Proposal of Haerentibaculum gen. nov. with Haerentibaculum
RT maritimum comb. nov. and Haerentibaculum ovolyticus comb. nov., and
RT two new species *;
RL Submitted (SEP 1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA
CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPISOMERASE FAMILY.

DR EMBL: AB032579; BAB13317.1;
DR HSSP: P06982; 1A36;
DR InterPro: IPR001241; DNA_topoisom.
DR InterPro: IPR003594; HATPase_c.
DR InterPro: IPR002936; Toprim.
DR Pfam: PF00204; DNA_topoisom_I; 1.
DR Pfam: PF02518; HATPase_c;
DR Pfam: PF01751; Toprim; 1.
DR PRINTS: PR00418; TP12FAMILY.
DR PRODOM: P0000616; DNA_topo_soi1; 1.
DR SMART: SM00433; TOP2c; 1.
DR PROSITE: PS00177; TOPISOMERASE_II; 1.
KW ATP-binding; Isomerase; Topoisomerase.
FI NON_TER 1
FI NON_TER 481
SQ SEQUENCE 481 AA: 53556 MW: 8707PFA3080A70H1 CRC64;

Query Match 4.1%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SENDPEKC 39
IIIIIIII
316 SENDPEKC 323

RESULT 5
Q9FAX0
ID Q9FAX0 PRELIMINARY: PRI: 481 AA.
AC Q9FAX0
DT 01-MAR-2001 (TREMREL: 16, Created)
DI 01-MAR-2001 (TREMREL: 16, Last sequence update)
DE 01-DEC-2001 (TREMREL: 19, Last annotation update)
DE DNA GYRASE B SUBUNIT (FRAGMENT).
GN GYRB.
OS Flexibacter illiformis.
OC Bacteria: CFB group: Flexibacter group: Flexibacter.
OX NCBI_TaxID=104663;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IFO 15056;
RA Suzuki M., Yamaguchi K.;
PT *Phylogenetic analysis and taxonomic study of marine Cytophaga like
RI bacteria. Proposal of Haerentibaculum gen. nov. with Haerentibaculum

Query Match 4.1%; Score 8; DB 7; Length 253;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 145 ALITVIAL 142

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein protein search, using sw model

Run on: July 2, 2002, 16:18:32, Search time 96.19 Seconds

(without alignments)
350.702 Million cell updates/sec

Title: 09-603-665-5_COPY_1950_2144

Perfect score: 195

Sequence: 1 LPTTFACHIVAKPRADMLXQV CQKTLQGLTVLGGHPQSYF 195

Scoring table:

Gapop 60 0, Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size: 6

Total number of hits satisfying chosen parameters: 1258

Minimum DB seq length: 0

Maximum DB seq length: 2060000000

Post-processing: Listing first 45 summaries

Database:

SUPREMH1_19:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_ordancelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	127	65	1	449	4	Q96ES5 homo sapien
2	8	4	1	188	16	Q9K6G3 bacillus ha
3	8	4	1	253	7	Q9GJN1 calman croc
4	8	4	1	481	2	Q9FAX2 flexibacter
5	8	4	1	481	2	Q9FAX0 flexibacter
6	8	4	1	481	2	Q9FAX9 flexibacter
7	8	4	1	481	2	Q9AOL4 chlinophag
8	7	3	6	72	2	Q44537 azolobacter
9	7	3	6	165	10	Q91ZV5 arabidopsis
10	7	3	6	205	4	Q90319 homo sapien
11	7	3	6	214	11	Q9D0P3 mus musculu
12	7	3	6	228	11	Q70473 cricetus
13	7	3	6	232	13	Q90WT1 brachydanio
14	7	3	6	238	2	Q9JPE8 neisseria m
15	7	3	6	252	11	Q9CT53 mus musculu
16	7	3	6	252	11	Q9CQ15 mus musculu

17	7	3	6	265	16	Q98HM1
18	7	3	6	285	6	Q9N045
19	7	3	6	296	16	Q9AAU6
20	7	3	6	305	2	Q93SV5
21	7	3	6	317	5	Q91020
22	7	3	6	325	11	Q9J116
23	7	3	6	325	11	Q9D017
24	7	3	6	325	11	Q9D016
25	7	3	6	325	11	Q9D012
26	7	3	6	325	11	Q9C018
27	7	3	6	326	16	Q55684
28	7	3	6	332	10	Q9FK92
29	7	3	6	333	13	Q9DGL7
30	7	3	6	336	2	Q9Z3C1
31	7	3	6	337	2	Q9E968
32	7	3	6	337	2	Q95738
33	7	3	6	337	5	Q45326
34	7	3	6	342	4	Q95C28
35	7	3	6	343	11	Q9CW04
36	7	3	6	343	11	Q99JR9
37	7	3	6	353	17	Q28890
38	7	3	6	359	2	Q88036
39	7	3	6	359	5	Q21008
40	7	3	6	374	5	Q9V313
41	7	3	6	374	10	Q9C502
42	7	3	6	391	2	Q952B3
43	7	3	6	393	16	Q9KEQ0
44	7	3	6	400	16	Q9KSB9
45	7	3	6	419	16	Q9KB36

ALIGNMENTS

RESULT 1

Q96FS5 ID Q96ES5 PRELIMINARY; PRT; 349 AA.

AC Q96FS5; Q96ES5; (TRENBLREL. 19, Created)

DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)

DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)

DE SIMILAR TO HYPOTHETICAL PROTEIN FLJ10359.

OS Homo sapiens (Human).

OC Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi;

OC Mammalia, Eutheria, Primates, Catarrhini, Homidae, Homo.

OX NCBI_TaxID:9606;

RN [1]

RP SOURCE FROM N.A.

RC TISSUE-OVARY, AND ADENOCARCINOMA;

RA Strausberg R.

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC011983; AHH1983.1; -;

SQ SEQUENCE 349 AA; 35921 MW; 3A359597FF7079EB CRC64;

Query Match 65.1%, Score 127, DR 4, Length 349;
Best Local Similarity 100.0%; Pred. No. 3e-122;
Matches 127, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 69 ALMPFLVQLLENRIAGGKFKVERIKHLIVCIAGFSVAMADDSWKPIKYNQIIKTRISS 128

|||||

DB 223 ALMPFLVQLLENRIAGGKFKVERIKHLIVCIAGFSVAMADDSWKPIKYNQIIKTRISE 282

QY 129 PKVFAALITVLAAELKENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLG 188

|||||

DB 283 PKVFAALITVLAAELKENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLG 342

QY 189 EPLQSYF 195

|||||

DB 343 EPLQSYF 349

RESULT 2


```

RESULTS: 15
V190_HELPY
ID V190_HELPY STANDARD: PRI: 502 AA.
AC P56117.
DT 01 NOV 1997 (rel. 35, created)
DI 01 NOV 1997 (rel. 35, last sequence update)
DI 01 MAR 2002 (rel. 41, last annotation update)
DE Hypothetical protein 490190.
GN HP0190
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI TaxID: 210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700492.
RX MEDLINE:9704467; PubMed:9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gil S., Dougherty H.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathley L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN
CC SYNTHASE SUBFAMILY. STRONG. TO E-COLI YMDC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sth.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000539; AAC07257.1;
DR U08: HP0190;
DR InterPro: IPR001736; 110.
DR Pfam: PF00614; Pfam: 2.
DR SMART: SM00155; Pfam: 2.
KW Hypothetical protein; Transferase; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
SQ SEQUENCE: 502 AA; 58287 MW; C83911F2F59C714 CRC64;

```

```

Query Match 4.6% Score 7; DB 1; Length 502;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 143 AEKLEKN 149
ID 44 AEKLEKN 50

```

```

Search completed: July 2, 2002, 16:19:09
Job time: 781 sec

```

DR InterPro: IPR000567; SBP_bac_1;
 DR Pfam: PF01547; SBP_bacterial_1; 1.
 DR PROSITE: PS0013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01037; SBP_BACTERIAL_1; FALSE_NEG.
 KW Hypothetical protein; Transport, Membrane, Lipoprotein; Signal;
 KW Complete proteome.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 441 HYPOTHEICAL ABC TRANSPORTER
 FT LIPID 24 24 EXTRACELLULAR BINDING PROTEIN PH1214.
 FT N-ACYL DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 441 AA; 4886 MW; 558P963D816CA7EE CRC64;

Query Match 3.6%; Score 7; DB 1; Length 441;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 AKIKKEN 149
 DB 191 AKIKKEN 197
 I I I I I I I

RESULT 13
 PURA_SYNY3 STANDARD; PRT; 444 AA.
 AC P73290;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP+aspartate ligase)
 DE (ADSS) (AMPSase).
 DE PURA OR SL1823.
 OS Synechocystis sp. (strain FGO 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN I I I I I
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima T., Hiroseawa M., Sudaira M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain FGO6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 FL DNA Res. 3:109-136(1996).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE
 CC NUCLEOTIDE BIOSYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate + phosphate +
 CC adenylosuccinate.
 CC -1- PATHWAY: FIRST COMMITTED STEP IN AMP BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE ADENYLOSUCCINATE SYNTHETASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 CC or send an email to license@sib-sib.ch).

EMBL: D90905; RAA17318.1; ALT_INIT.
 DR HSSP: P12283; IADE.
 DR InterPro: IPR001114; Adenylosucc_synth.
 DR Pfam: PF00709; Adenylosucc_synth; 1.
 DR ProDom: PD001188; Adenylosucc_synth; 1.
 DR PROSITE: PS00113; ADENYLOSUCCIN_SYN.2; 1.
 DR PROSITE: PS01266; ADENYLOSUCCIN_SYN.1; 1.
 KW Purine biosynthesis; Ligase; GTP-binding; Complete proteome.
 FT NP_BIND 13 19
 FT ACT_SITE 140 140 HY SIMILARITY.
 FT ACT_SITE 147 147 HY SIMILARITY.

SO SEQUENCE 444 AA; 48792 MW; F064900P0F241A4RF CRC64;

Query Match 3.6%; Score 7; DB 1; Length 463;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 LAEKLKE 148
 DB 86 LAEKLKE 92
 I I I I I I I

RESULT 14
 VDHA_CHICK STANDARD; PRT; 463 AA.
 AC Q90578;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Vitamin D3 hydroxylase-associated protein (VDHAP).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN I I I I I
 RP SEQUENCE FROM N.A.
 RX STRAIN=WHITE LECHORN; TISSUE=Kidney;
 RX MEDLINE=94103206; PubMed=8276793;
 RA Ptinger R.A., Ismail R., Deluca H.F.;
 RT "cDNA cloning and characterization of a vitamin D3 hydroxylase-
 RT associated protein.";
 FL J Biol. Chem. 269:176-182(1994)
 RN I I I I I
 RP SEQUENCE OF 1-20.
 RC STRAIN=WHITE LECHORN; TISSUE=Kidney;
 RX MEDLINE=92147714; PubMed=1310688;
 RX Burgos-Trinidad M., Ismail R., Ptinger R.A., Deluca H.F.;
 RT "Immunopurified 25-hydroxyvitamin D 1 alpha hydroxylase and 1,25-
 RT dihydroxyvitamin D 24-hydroxylase are closely related but distinct
 RT enzymes.";
 RL J. Biol. Chem. 267:3498-3505(1992).
 CC -1- FUNCTION: MAY HAVE A VITAMIN D3 HYDROXYLASE REGULATORY FUNCTION.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
 CC -1- TISSUE SPECIFICITY: KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 CC or send an email to license@sib-sib.ch).

EMBL: U00694; AAC59645.1; -.
 DR InterPro: IPR000120; Amidase.
 DR Pfam: PF01425; Amidase; 1.
 DR PROSITE: PS00571; AMIDASES; 1.
 KW Hydrolyase; Mitochondrion; Membrane.
 FT INIT_MET 0 0
 FT CONFLICT 5 6 LW -> PS (IN REF. 2).
 FT CONFLICT 13 13 W -> G (IN REF. 2).
 SQ SEQUENCE 463 AA; 50760 MW; 68E7F5F48FB06497 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 463;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 LAEKLKE 148
 DB 86 LAEKLKE 92
 I I I I I I I

1 SIMILARITY: RELATIONS TO THE ALDO/KETO REDUCTASE FAMILY.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D10854; AAA01627.1;
 DR HSSP: P14550; 2A8R.
 DR InterPro: IPR001195; Aldo_ket_red
 DR Pfam: PF00248; Aldo_ket_red; 1.
 DR PRINTS: PR00069; ALDKREDASE.
 DR PROSITE: PS00062; ALDKETO-REDUCTASE_2; 1.
 DR PROSITE: PS00063; ALDKETO-REDUCTASE_3; 1.
 DR PROSITE: PS00798; ALDKETO-REDUCTASE_1; 1.
 KW Oxidoreductase; NADP; Acetylation.
 FT INIT_MET 0 1 BY SIMILARITY.
 FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 FT ACET_SITE 112 112 HYDROGEN BOND DONOR (BY SIMILARITY).
 SQ SEQUENCE 424 AA; 36474 MW; A38C5EA6F0F9C5F2 CRC64;

Query Match 3.6% Score 7; DR 1; Length 424;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 VLAAEK 145
 I I I I I I
 Lb 244 VLAAEK 249

RESULT 11
 ADHA_RHIME
 ID ADHA_RHIME STANDARD; PRT: 340 AA.
 AC 041186;
 DI 30-MAY-2000 (Rel. 19; Created)
 DI 30-MAY-2000 (Rel. 19; Last sequence update)
 DI 01-MAR-2002 (Rel. 41; Last annotation update)
 DE Alcohol dehydrogenase (EC 1.1.1.1).
 GN ADHA OR KA0704 OR SMA1296.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 C2 Plasmid pSymA (megaplasmid 1).
 CC Bacteroid: Proteobacteria; alpha subdivision; Rhizobiaceae group.
 CC Rhizobiaceae; Sinorhizobium.
 CX NCBI_TaxID=482;
 RP SEQUENCE FROM N.A.
 RX STRAIN=1021;
 RA MEDLINE=98323559; PubMed=9659480;
 RA WILLIS L.H., Walker G.S.;
 RT "Identification of the Rhizobium meliloti alcohol dehydrogenase gene
 RT (adha) and heterologous expression in Alcaligenes eutrophus.";
 RL Biochim. Biophys. Acta 1384:197-203(1998).
 RN 12)
 RP SEQUENCE FROM N.A.
 RX STRAIN=1021;
 RA MEDLINE=21496509; PubMed=11481442;
 RA Barrett M.J., Fisher R.F., Jones T., Kemp G., Abola A.P.,
 RA Bartley Hubler F., Bowers L., Capela D., Galibert F., Gouzy J.,
 RA Gurjal M., Hong A., Holzer L., Hyman R.W., Kahn D., Kahn M.,
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Serrazek R., Wells D.H.,
 RA Yeh K.C., Davis R.W., Federapfel N.A., Long S.K.;
 RT "Nucleotide sequence and predicted functions of the entire
 RT Sinorhizobium meliloti pSymA megaplasmid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) -> an aldehyde or ketone
 CC NADH.
 CC -1- COFACTOR: ZINC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY. HIGH, WITH OTHER BACTERIAL ADH S.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF031940; AAR87463.1;
 DR EMBL: AF007258; AAK65362.1;
 DR HSSP: P00325; IDEH.
 DR InterPro: IPR003328; ADH_zinc.
 DR InterPro: IPR002085; Adh_z_family.
 DR Pfam: PF00107; adh_zinc; 1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; Zinc; NAD; Complete proteome.
 FT METAL 40 40 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 63 63 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 340 AA; 36235 MW; C1A04B5FEE5C95E1 CRC64;

Query Match 3.6% Score 7; DR 1; Length 340;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 LALAEKI 146
 I I I I I I
 Lb 203 LALAEKI 209

RESULT 12
 YC14_PYRHO
 ID YC14_PYRHO STANDARD; PRT: 441 AA.
 AC 058969;
 DI 15-DEC-1998 (Rel. 37; Created)
 DI 15-DEC-1998 (Rel. 37; Last sequence update)
 DI 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Hypothetical ABC transporter extracellular binding protein PH1214
 DE precursor.
 GN PH1214 OR PHBK040.
 OS Pyrococcus horikoshii.
 CC Archaea; Euryarchaeota; Thermococcales; Thermococcus.
 CX NCBI_TaxID=54953;
 RN 11)
 RP SEQUENCE FROM N.A.
 RX STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9479194;
 RA Kwarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baka S.-I., Kusui H., Hosoyama A., Naeai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Ouchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Nishikawa K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA RES. 5:55-76(1998).
 CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM PH1214/15/16.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (potential).
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 CC PROTEIN FAMILY 1.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AP000005; BAA30314.1;


```

FT DISULFID 118 138 RY SIMILARITY.
FT CONFLICT 114 115 EH -> OE (IN REF. 3).
SQ SEQUENCE 141 AA: 15949 MW: D7632905541CB266 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 141;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 ENYIVLL 154
DB 128 ENYIVLL 134

RESULT 4
RL9 MYCGE STANDARD; PRT; 150 AA.
AC P47339;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L9.
GN RPL9 OR RPL9 QM MG093.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmatellaceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN 111
SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569903;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Holt C.F., Korfavage A.P., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Tomb J.-F., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995)
CC 1- FUNCTION: BINDS TO THE 23S RRNA (BY SIMILARITY).
CC 1- SIMILARITY: BELONGS TO THE L9P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL: U39689; AAC71311.1; -.
DR HSSP; P02417; 1b1v.
DR TIGR; MG093; -.
DR InterPro; IPR000244; Ribosomal_L9.
DR Pfam; PF01281; Ribosomal_L9; 1.
DR PROSITE; PS00651; RIBOSOMAL_L9; FALSE NEG.
KW Ribosomal protein; rRNA-binding, Complete proteome.
SQ SEQUENCE 150 AA: 17380 MW: 50A19B6EAA48F8D6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 150;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 ALAEKIK 147
DB 63 ALAEKIK 69

RESULT 5
F3GI_ADR05 STANDARD; PRT; 160 AA.
AC P04494;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 160;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 LITVIAL 142
DB 132 LITVIAL 138

RESULT 6
KIDS_HELPUJ STANDARD; PRT; 276 AA.
AC Q92N55;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)

```

```

FT VARIANT 2017 2017 /FTIG-VAR_0941.
FT E E > G.
FT /FTIG-VAR_0942.
SQ SEQUENCE 2144 AA: 212455 MW: D66816E78D8C9H7 CRC64:

Query Match
Best Local Similarity 100.0%; Pred. No. 64; Length 2144;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALMPPLVDGLENPLGGEKPEFQVTVTHLINC:AGISVAMADRSWKTVNYQ:HLKTRSS 128
      |||||
DB 2018 ALMPPLVDGLENPLGGEKPEFQVTVTHLIPC:AGISVAMADRSWKPLNYQILLTRSS 2077
      |||||

QY 129 PKVRFPAALIVLALAEKIKENYIVLLPES:PF:ALMEDECEVEHQCOKTQQO:ETVL 188
      |||||
DB 2078 PKVRFPAALIVLALAEKIKENYIVLLPES:PF:ALMEDECEVEHQCOKTQQO:ETVL 2147
      |||||

QY 189 EPLOSYP 155
      |||||
DB 2148 EPLOSYP 2144
      |||||

RESULT 2
BE28 MACFA STANDARD: PRT: 98 AA.
AC Q9GMA44:
DT 16 OCT-2001 (Ref. 40, Created)
DI 16 OCT-2001 (Ref. 40, Last sequence update)
DE Protein BAP28 (Fragment).
GN BAP28
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Pukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP TISSUE=Brain;
RC TISSUE=Brain;
RA Osada N., Hida M., Kusuda J., Tanuma K., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sudano S., Iashimizu K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BAP28 FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HEAT REPEAT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/submit/
CC or send an email to license@ebi.ac.uk).
DB EMBL: AH049842; BAB16728.1; ALT_INIT.
DR InterPro: IPR000457; HEAT_REPEAT.
DR PROSITE: PS00077; HEAT_REPEAT; FALSE_NEG.
FT NON TER 1
FT REPEAT 920 956 HEAT.
SQ SEQUENCE 956 AA: 104644 MW: 3DBD95C3623CFB31 CRC64:

Query Match
Best Local Similarity 100.0%; Pred. No. 64; Length 956.
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 SPKVRFAALIVLALAEKIKENYIVLLPES:PF:ALMEDECEVEHQCOKTQQO:ETVL 187
      |||||
DB 891 SPKVRFAALIVLALAEKIKENYIVLLPES:PF:ALMEDECEVEHQCOKTQQO:ETVL 950
      |||||

QY 188 GEPL 191

```

```

DB 951 GEPL 954
      |||||

RESULT 3
CYTS_RAT
ID CYTS_RAT STANDARD: PPT: 141 AA.
AC P19313:
DT 01-NOV-1990 (Ref. 16, Created)
DI 01-FEB-1996 (Ref. 33, Last sequence update)
DI 01-MAR-2002 (Ref. 41, Last annotation update)
DE Cystatin S precursor (LM protein).
GN CST4 OR CYSS.
OS Rattus norvegicus (Rat).
OC Pukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92165056; PubMed=1537554;
RA Cox J.L., Shaw P.A.;
RT "Structure, organization and regulation of a rat cysteine proteinase
RT inhibitor-encoding gene."
RL Gene 110:175-180(1992).
RN [2]
RP SEQUENCE OF 10-141 FROM N.A.
RC STRAIN=SPRACUE-DAWLEY; TISSUE=Submandibular gland;
RX MEDLINE=89053983; PubMed=3263967;
RA Shaw P.A., Cox J.L., Harka T., Naito Y.;
RT "Cloning and sequencing of cDNA encoding a rat salivary cysteine
RT proteinase inhibitor inducible by beta adrenergic agonists."
RL J Biol Chem 263:18133-18137(1988).
RN [3]
RP SEQUENCE OF 28-132.
RC TISSUE=Submandibular gland;
RX MEDLINE=89334379; PubMed=2757396;
RA Bedi G.S.;
RT "Amino acid sequence of an inducible cysteine proteinase inhibitor
RT (cystatin) from submandibular glands of isoproterenol-treated rats."
RL Arch. Biochem. Biophys. 273:245-253(1989).
CC -!- FUNCTION: THIS PROTEIN STRONGLY INHIBITS PAPAIN & FICIN, PARTIALLY
CC INHIBITS STEM BRONELAIN & BOVINE CATHEPSIN C, BUT DOES NOT INHIBIT
CC PAPINE CATHEPSIN B OR CLOSTROPAIN. PAPAIN IS INHIBITED
CC NONCOMPETITIVELY.
CC -!- TISSUE SPECIFICITY: FOUND IN SALIVA, TEARS, URINE, AND SEMINAL
CC FLUID.
CC -!- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/submit/
CC or send an email to license@ebi.ac.uk).
DB EMBL: M75281; AAA41068.1;
DR EMBL: J04206; AAB59703.1;
DR PIR: A31891; A31891
DR PIR: S05252; S05252.
DR HSSP: P01038; ICEW.
DR InterPro: IPR000010; Cystatin.
DR InterPro: IPR003243; Cystatin_C_M.
DR Pfam: PF00041; Cystatin; 1.
DR ProDom: PD001231; Cystatin_C_M; 1.
DR SMART: SM00043; CY; 1.
DR PROSITE: PS00287; CYSTATIN; 1.
KW Thiol protease inhibitor; Signal.
FT SIGNAL 1 27
FT CHAIN 28 141 CYSTATIN S.
FT ACT_SITE 32 32 REACTIVE SITE.
FT SITE 76 80 SECONDARY AREA OF CONTACT.
FT DISULFID 94 104 BY SIMILARITY.

```

genforce version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:19:08 : Search time 30.19 seconds
(without alignments)
250.093 Million cell updates/sec

Title: US-09 603 665-5_COPY_1950_2144

Perfect score: 195

Sequence: 1 LFTLFGHLVRPFALTLQGV.....LCKTIQQLVILGPELQSYF 195

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size: 6

Total number of hits satisfying chosen parameters: 340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	127	65.1	2144	HP28_HUMAN	Q9H583 homo sapien
2	64	32.8	968	BP28_MACP7	Q9QM44 macaca fasc
3	7	3.6	141	CYTS_PAT	P19313 rattus norv
4	7	3.6	150	RL9_MYGE	P47339 mycoplasma
5	7	3.6	160	1_EGLI_ADRO	P04494 human adeno
6	7	3.6	276	KD5A_HELP	Q92655 helicobacte
7	7	3.6	307	QIC_ARCF	G29013 archaebaclob
8	7	3.6	316	MCNA_PYRIB	Q57854 pyrococcus
9	7	3.6	324	ALDX_HUMAN	P14550 homo sapien
10	7	3.6	324	ALDX_RAT	P51635 rattus norv
11	7	3.6	340	ADHA_RHIM6	Q31186 rhizobium m
12	7	3.6	441	YGL4_PYRHO	Q58969 pyrococcus
13	7	3.6	444	PURA_SYNY3	P73290 synecocyst
14	7	3.6	463	VDHA_CHICK	Q90578 gallus galli
15	7	3.6	502	Y190_HELPY	P56117 helicobacte
16	7	3.6	556	PIR_CANAU	P34205 carassius a
17	7	3.6	731	CATA_HAUMA	Q59651 haloraula
18	7	3.6	885	CHS3_EXODE	P36062 exophiala d
19	7	3.6	911	CHSG_ASFPU	P54267 aspergillus
20	7	3.6	916	CHSG_EMENI	Q00757 americella
21	7	3.6	1045	IMD2_YEAST	P12684 saccharomyc
22	7	3.6	1121	DPX8_APATH	Q38953 arabidopsis
23	7	3.6	1122	ADP1_MYCCA	Q49379 mycoplasma
24	7	3.6	1830	HP28_APATH	Q9CH24 arabidopsis
25	6	3.1	73	RL7_STAAR	P48620 staphylococ
26	6	3.1	87	GLRX_HAEIN	P45242 haemophilus
27	6	3.1	96	REV_SIVAM	P36339 simian immu
28	6	3.1	101	FEV_SIVGB	P23479 simian immu
29	6	3.1	113	RU24_FUNHU	P33134 micrococcus
30	6	3.1	114	GTHI_FUNHE	P30971 funulus he
31	6	3.1	118	THIH_RICCO	Q43636 ricinus com
32	6	3.1	125	MERR_STREI	P30346 streptomyce
33	6	3.1	132	YVGA_VAGCG	P20515 vaccinia vi

ALIGNMENTS

RESULT 1

ID	BP28_HUMAN	STANDARD:	PRT:	2144 AA.
AC	Q9H583; Q9NW23;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Protein BAP28.			
GN	RAP28.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANTS S-1694; A-1854; D-1967 AND G-2017.			
RA	Houquelet L., Chumakov I., Barry C., Cohen Aknine A.,			
RT	"A novel HAP28 gene and protein."			
RL	Patent number WO0100669, 04-JAN-2001.			
RN	[2]			
RP	SEQUENCE OF 1534-2144 FROM N.A.			
RA	Cobley V.,			
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1777-2144 FROM N.A.			
RA	Isoqai T., Ota T., Hayashi K., Sugiama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sudano S., Shiratori A., Sudo H.,			
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y.,			
RA	Ninomiya K., Iwayanagi T.,			
RT	"NEDO human cDNA sequencing project."			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE HAP28 FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 HEAT REPEAT.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL: AX067150; CA26776.1; ..			
DR	EMBL: A136105; CAC15948.1; ..			
DR	EMBL: AK001221; HAA91564.1; A1T_INIT.			
DR	InterPro: IPR000357; HEAT_repeat.			
DR	PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.			
KW	Polymorphism			
FT	REPEAT 2106 2142 HEAT.			
FT	VARIANT 1694 1694 N -> S.			
FT	VARIANT 1854 1854 V -> A.			
FT	VARIANT 1967 1967 N -> D.			

P45928 bacillus su
P15145 human adeno
C97159 pyrococcus
P42180 pyrococcus
Q59256 pyrococcus
P14527 oncorhynchu
P48530 xenopus lae
Q42425 gadus morhua
P16418 liophis mil
P54057 methanococ
Q9JWV7 neisseria m
Q9JXZ9 neisseria m

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

Qy 142 LAEKLEK 148
| | | | |
Db 105 LAEKLEK 111

Search completed. July 2, 2002, 16:15:42
Job time: 759 sec

C: Mazur, M.; Galtisano, E.; Seikawa, F.; Elzer, P.H.; Haddad, S.; O'Fallon, B.; Levesque, N.; Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen *Bruceella melitensis*
 A:Reference number: AB1252; PMID:11756688

A:Accession: AG1405
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-276 <KUR>
 A:Cross-references: GB:AF008917; PIDN:AA151610.1; PID:q17982446; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BME10429
 A:Map position: 1
 C:Keywords: methyltransferase

Query Match 3.6%; Score 7; DB 2; Length 276;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 78 LENRIAG 84
 Db 67 LENRIAG 73

RESULT 12
 A10190
 Probable hydrolase (EC 5.3.3.-) [imported] - *Yersinia pestis* (strain CO92)
 C:Species: *Yersinia pestis*
 C:Date: 02 Nov 2001 #sequence_revision 02 Nov 2001 #ext_change 14 Feb 2002
 C:Accession: A10190
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.R.; de Lencastre, A.M.; Chittliffe, L.; Cronin, A.; Davies, P.M.; Davis, P.; Dougan, G.; et al.; M.J. Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10190
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-280 <KUR>
 A:Cross-references: GB:AL550842; PIDN:CAC90488.1; PID:q15979606; GSPDB:GN00175
 C:Genetics:
 A:Gene: YP01566
 C:Superfamily: 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase
 C:Keywords: intramolecular oxidoreductase; isomerase

Query Match 3.6%; Score 7; DB 2; Length 280;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 175 QGQRTIQ 181
 Db 273 QGQRTIQ 279

RESULT 13
 A87311
 Hypothetical protein G0498 [imported] - *Caulobacter crescentus*
 C:Species: *Caulobacter crescentus*
 C:Date: 20 Apr 2001 #sequence_revision 20 Apr 2001 #ext_change 20 Apr 2001
 C:Accession: A87311
 R:Norman, W.C.; Foldblyum, T.V.; Paulsen, I.F.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; Laub, M.; DeWay, K.L.; Doolan, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Kolon, J.; Emulaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4146-4141, 2001
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: A87311
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-296 <STO>

A:Cross-references: GB:AE005673; NID:q1441477; PIDN:AA12495.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0498

Query Match 3.6%; Score 7; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LFTLFAG 7
 Db 240 LFTLFAG 246

RESULT 14
 F69406
 Ornithine carbamoyltransferase (argP) homolog - *Archaeoglobus fulgidus*
 C:Species: *Archaeoglobus fulgidus*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec 1997 #ext_change 18 Jun 1999
 C:Accession: F69406
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Led, G.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sarr, G.; Gill, S.; Kirkness, E.; et al.; Fleischmann, R.D.; Overback, R.; Gocayne, J.D.; Weidman, J.F.; McLeod, L.; Glodek, A.; Zhou, L.; Nature 390, 364-370, 1997
 A:Authors: Overback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch.
 A:Reference number: A59250; MUID:9804943
 A:Accession: F69406
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-307 <KLE>
 A:Cross-references: GB:AF031617; GB:AF100782; NID:q289440; PIDN:AA85987.1; PID:q264
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase; F2-300/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 3.6%; Score 7; DB 2; Length 307;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 142 LAEKLKE 148
 Db 21 LAEKLKE 27

RESULT 15
 H71231
 molybdopterin biosynthesis protein moaA - *Pyrococcus horikoshii*
 C:Species: *Pyrococcus horikoshii*
 C:Date: 14 Aug 1998 #sequence_revision 14-Aug-1998 #ext_change 24 Oct 2000
 C:Accession: H71231
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; So, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Koshida, N.; Oda, DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper thermophilic
 A:Reference number: A71000; MUID:98344137
 A:Accession: H71231
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-316 <KAW>
 A:Cross-references: GB:AF009301; NID:q246128; PIDN:BAW5186.1; PID:q156500
 A:Experimental source: strain 093
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH0114
 C:Superfamily: Escherichia coli molybdopterin biosynthesis protein moaA
 C:Keywords: molybdopterin biosynthesis

Query Match 3.6%; Score 7; DB 2; Length 316;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

J. Biol. Chem. 260, 2424-2431, 1985
 A:Title: The 19-kDa glycoprotein coded by region E3 of adenovirus.
 A:Reference number: A22515; PMID:85130985
 A:Accession: A22515
 A:Molecule type: DNA
 A:Residues: 1-160 <WOL>
 A:Cross-references: GR:M12406; NID:q200011; PDB:AAA42492.1; PDB:q200012
 C:Superfamily: adenovirus early E3 18.5K glycoprotein
 C:Keywords: early protein; glycoprotein; transmembrane protein
 E:30/79/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.6%, Score 7, DB 1, Length 160,
 Best Local Similarity 100.0%; Pred. No. 21,
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 LITVLAL 142
 DQ 132 LITVLAL 138
 |||||

RESULT 7
 T48212
 hypothetical protein T20L15.180 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse ear cress)
 C:Date: 20 Apr 2000 #sequence_revision 20 Apr 2000 #text_change 20 Apr 2000
 C:Accession: T48212
 P:Bevan, M., Peters, S.A., van Slaveren, M., Dirksen, W., Stiekema, W.; Bancroft, I.; Mow
 submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24488
 A:Accession: T48212
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-165 <BEV>
 A:Cross-references: EMBL:AL162351
 A:Experimental source: cultivar Columbia; BAC clone T20L15
 C:Genetics:

A:Map position: 5
 A:Introns: 99/3; 121/3
 A:Note: T20L15.180

Query Match 3.6%, Score 7, DB 2, Length 165,
 Best Local Similarity 100.0%; Pred. No. 21,
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 AEKLEKN 149
 DQ 30 AEKLEKN 36
 |||||

RESULT 8
 T14745
 hypothetical protein DKFZp586G1624.1 human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14745
 R:Gasterbauer, A., Lauber, J., Meers, H.W.; Gasterbauer, J.; Wiemann, S.
 submitted to the Protein Sequence Database, August 1999
 A:Reference number: Z18178
 A:Accession: T14745
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-205 <DUE>
 A:Cross-references: EMBL:AL110160
 A:Experimental source: adult uterus; clone DKFZp586G1624
 C:Genetics:
 A:Note: DKFZp586G1624.1

Query Match 3.6%, Score 7, DB 2, Length 205,
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TVLGEPL 191
 DQ 137 TVLGEPL 143
 |||||

RESULT 9
 JC7706
 beta crystallin B1 protein - zebra fish
 C:Species: Brachydanio rerio (zebra fish)
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: JC7706
 P:Chen, J.Y., Chang, R.F.; Chen, Y.H.; Lin, C.J.F.; Wu, J.T.; Kuo, C.M.
 Biochem Biophys Res Commun 285, 104-110, 2001
 A:Title: Molecular cloning, developmental expression, and hormonal regulation of zeb
 A:Reference number: JC7706; MUID:21331298; PMID:11437379
 A:Accession: JC7706
 A:Molecule type: mRNA
 A:Residues: 1-232 <CHE>
 A:Cross-references: GR:AJ317957
 C:Comment: This protein is involved in regulation by growth factors.
 C:Superfamily: beta crystallin
 E:20-25/h-region: active proline and alanine-rich motif #status predicted

Query Match 3.6%, Score 7, DB 2, Length 232,
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 YKIFLFD 57
 DQ 41 YKIFLFD 47
 |||||

RESULT 10
 E71985
 3-deoxy-D-manno-octulosonic acid 8-phosphate synthase - Helicobacter pylori (strain J
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: E71985
 R:Alm, R.A.; Ling, L.S.; Moir, D.T.; King, H.L.; Brown, E.D.; Doid, P.C.; Smith, D.
 J. J. C. Gibson, P. J. Metberg, D. J. Mills, S. D. Jiang, Q. J. Taylor, D. E. J. Vovis, G. F.
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
 A:Reference number: A71800; MUID:99120557
 A:Accession: E71985
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-276 <ARN>
 A:Cross-references: GR:AE001449; GR:AE001450; NID:q4154501; PDB:AAU05587.1; PID:q415
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: kdsA
 C:Superfamily: phospho-2-dehydro 3 deoxyoctonate aldolase

Query Match 3.6%, Score 7, DB 2, Length 276,
 Best Local Similarity 100.0%; Pred. No. 33,
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 LKTRDSS 128
 DQ 148 LKTRDSS 154
 |||||

RESULT 11
 AG3305
 23S ribosomal PNA methyltransferase (EC 2.1.1.-) [imported] - Brucella melitensis (st
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AG3305
 R:Belvecchio, V.G.; Kaputrat, V.; Pedkar, P.J.; Dutra, C.; Mijer, C.; Los, T.; Ivanov

Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 87 KFORV 93
 DB 63 KFORV 69
 |||||

RESULT 4
 Cystatin S precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #extl_change 16-Jul-1999
 C:Accession: J01470; A13991; S05252; S15132
 F: Cox, J.L.; Shaw, P.A.
 F: Gene 110, 175-180, 1992
 A: Title: Structure, organization and regulation of a rat cysteine proteinase inhibitor-c
 A: Reference number: J01470; M010: 92165056
 A: Accession: J01470
 A: Molecule type: DNA
 A: Residues: 1-141 <SHA>
 A: Cross-references: GH: M75281; NID: q244537; PIDN: AAA41068.1; PID: g294538
 F: Shaw, P.A.; Cox, J.L.; Batka, I.; Naito, Y.
 F: J. Biol. Chem. 263, 18133-18137, 1988
 A: Title: Cloning and sequencing of cDNA encoding a rat salivary cysteine proteinase inh
 A: Reference number: A31891; M010: 89054983
 A: Accession: A31891
 A: Molecule type: mRNA
 A: Residues: 175-254
 F: Hedi, G.S.
 A: Title: Amino acid sequence of an inducible cysteine proteinase inhibitor (cystatin) fr
 A: Reference number: S05252; M010: 8934379
 A: Accession: S05252
 A: Molecule type: protein
 A: Residues: 28-113, 116-141 <BED>
 A: Note: 87-Glu and 88-His were also found; disulfide bonds determined
 F: Nishikura, T.; Ishibashi, K.; Abe, K.
 F: Biochim. Biophys. Acta 1077, 346-354, 1991
 A: Title: Isolation of three forms of cystatin from submandibular saliva of isoproterenol
 A: Reference number: S15142; M010: 91240145
 A: Accession: S15142
 A: Molecule type: protein
 A: Residues: 29-48, 76-80 <BIO>
 A: Note: forms BSO-1, BSO-2 and BSO-3 with differing amino-terminals were found; a form B
 C: Genetics:
 A: Gene: CYSB
 A: Introns: 76/77, 114/73
 C: Superfamily: cystatin; cystatin homology
 C: Keywords: cysteine proteinase inhibitor
 F: 128/141/Domain: signal sequence #status predicted <SIG>
 F: 129-141/Product: cystatin S #status experimental <MAT>
 F: 130-141/Domain: cystatin S, form BSO-3 #status predicted <MAT3>
 F: 141/Product: cystatin S, form BSO-2 #status predicted <MAT2>
 F: 142-141/Product: cystatin S, form BSO-1 #status predicted <MAT1>
 F: 94-104, 118-138/Disulfide bonds: #status experimental

Query Match 3.6% Score 7: DB 2: Length 141:
 Best Local Similarity 100.0% Pred. No. 19:
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 148 ENYIVL 154
 DB 128 ENYIVL 134
 |||||

RESULT 4
 Ribosomal protein L9 - Mycoplasma genitalium
 C:Species: Mycoplasma genitalium
 C:Date: 17 Nov 1995 #sequence_revision 17 Nov 1999 #extl_change 30-Feb-1999
 C:Accession: G64210

ReFraser, C.M.; Gervayne, T.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.
 M.; Fuhrmann, J.; Nguyen, P.; Nierbach, T.P.; Saudek, D.M.; Phillips, G.A.; Merrick,
 C.A.; Venter, J.C.
 Science 270, 397-403, 1995
 A: Title: The minimal gene complement of Mycoplasma genitalium.
 A: Reference number: A64200; M010: 96026346
 A: Accession: G64210
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-150 <TIGR>
 A: Cross-references: GH: U99688; GH: L43967; NID: g1045763; PID: g1045770; TIGR: M6094
 A: Experimental source: strain G-37
 C: Genetics:
 A: Genetic code: SGC3
 C: Superfamily: Escherichia coli ribosomal protein L9

Query Match 3.6% Score 7: DB 2: Length 150:
 Best Local Similarity 100.0% Pred. No. 20:
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 141 ALAEKIK 147
 DB 63 ALAEKIK 69
 |||||

RESULT 5
 AC3385
 Retrovirus related pol polyprotein [imported] - Brucella melitensis (strain 16M)
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #extl_change 01-Feb-2002
 C:Accession: AC3385
 F: Belvecherio, V.G.; Kapral, V.; Rajkovic, P.J.; Patta, G.; Mujic, G.; Los, T.; Ivanov
 F: Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A: Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A: Reference number: AD3252; PMID: 11756688
 A: Accession: AC3385
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-151 <KUR>
 A: Cross-references: GH: AE008917; PIDN: AAL52246.1; PID: q17983042; GSPDB: GN00190
 A: Experimental source: strain 16M
 C: Genetics:
 A: Gene: HME11065
 A: Map position: 1

Query Match 3.6% Score 7: DB 2: Length 151:
 Best Local Similarity 100.0% Pred. No. 20:
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 135 ALITVIA 141
 DB 23 ALITVIA 29
 |||||

RESULT 6
 EBA45
 Early E3 18.5K glycoprotein - human adenovirus 5
 C:Species: Mastadenovirus h5 (human adenovirus 5)
 C: Note: host Homo sapiens (man)
 C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #extl_change 16-Jul-1999
 C:Accession: A03822; A22515
 F: Cladaras, C.; Wold, W.S.M.
 Virology 140, 28-43, 1985
 A: Title: DNA sequence of the early E3 transcription unit of adenovirus 5.
 A: Reference number: A94335; M010: 85092388
 A: Accession: A03822
 A: Molecule type: DNA
 A: Residues: 1-160 <CLA>
 A: Cross-references: GH: X00023; NC: 056503; EMBL: X00023; FID: 45806
 F: Wold, W.S.M.; Cladaras, C.; Deutscher, S.L.; Rapoport, G.S.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:15:43, Search time: 57.09 seconds
(without alignments)

32M 20M Million cell updates/sec

Title: US 09 603 665 5_COPY_1950_2144

Perfect score: 195

Sequence: 1 LPTTFAGHVKVPRAPRLXGV.....LQKFIQLLELVIGRPLQSYF 195

Scoring table:

OLIGO

Gapop 60.0, Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size: 6

Total number of hits satisfying chosen parameters: 931

Minimum DH seq length: 0

Maximum DH seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

PIR71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DH	ID	Description
1	8	4.1	188	2	F84120	hypothetical prote
2	7	3.6	98	2	S26914	Ig heavy chain V r
3	7	3.6	141	2	QJ1470	cystatin S precurs
4	7	3.6	150	2	C64210	ribosomal protein
5	7	3.6	151	2	AC3385	retrovirus-related
6	7	3.6	160	1	PRADA5	early E3 18.5K gly
7	7	3.6	165	2	T48212	hypothetical prote
8	7	3.6	205	2	T14745	hypothetical prote
9	7	3.6	232	2	JC7706	beta crystallin B1
10	7	3.6	276	2	F71985	3-deoxy-d-manno-oc
11	7	3.6	276	2	AC3305	23S ribosomal RNA
12	7	3.6	280	2	A01190	probable hydrolase
13	7	3.6	296	2	A87311	hypothetical prote
14	7	3.6	307	2	F69406	ornithine carbamoy
15	7	3.6	316	2	H71231	myeloperoxidase
16	7	3.6	317	2	T25526	hypothetical prote
17	7	3.6	325	1	JH0629	alcohol dehydrogen
18	7	3.6	325	2	A33851	alcohol dehydrogen
19	7	3.6	326	2	S76360	hypothetical prote
20	7	3.6	337	2	T30F44	hypothetical prote
21	7	3.6	340	2	H95349	alcohol dehydrogen
22	7	3.6	343	2	U69472	420-nonreducing h
23	7	3.6	359	2	T35179	oxidoreductase alp
24	7	3.6	359	2	T22950	hypothetical prote
25	7	3.6	374	2	R86457	unknown protein, 1
26	7	3.6	391	2	T36739	hypothetical prote
27	7	3.6	393	2	G83749	methionine gamma l
28	7	3.6	400	2	C94213	conserved hypothet
29	7	3.6	415	2	S77471	adenylosuccinate s

ALIGNMENTS

RESULT 1

F84120

hypothetical protein BH3766 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01 Dec 2000 #sequence_revision 01 Dec 2000 #text_change 15-Jun-2001

C:Accession: F84120

E:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, K.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A81650; MIMD:20512592; PMID:11058133

A:Accession: F84120

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-188 <STO>

A:Cross references: GH A001519, GH HA006034, NIDG10176103, PIRN:HA07485.1; GSPDR:G

A:Experimental source: Strain C-125

C:Genetics:

A:Gene: BH3766

C:Superfamily: Bacillus subtilis hypothetical protein ywlc

Query Match 4.1%; Score 8; DH 2; Length 188;

Best Local Similarity 100.0%; Prod. No. 2.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 QQLETVLG 188

|||||||

Db 6 QQLETVLG 13

RESULT 2

S26914

Ig heavy chain V region (DH-2) - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1993 #text_change 21 Jan 2000

C:Accession: S26914

P:Tomlinson, I.M.; Waller, G.; Marks, J.D.; Ilwellyn, M.B.; Winter, G.

J. Mol. Biol. 227, 776-798, 1992

A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups o

A:Reference number: S26985; MIMD:9302117

A:Accession: S26914

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-98 <TOU>

A:Cross references: EMBL:J2304; NIDG32963; PIRN:CAA78174.1; PIR:G3284

C:Superfamily: immunoglobulin V region, immunoglobulin homology

C:Keywords: heterotrimer; immunoglobulin

F.1.5.98/2domain, immunoglobulin homology -IMM-

Query Match

Best Local Similarity

4.5%; Score 7; DH 2; Length 98;

100.0%; Prod. No. 14;

DB 82 LALA-kl 88
|||||

RESULT 15

AAM69827 standard; Protein; 110 AA.

XX AAM69827;

XX 06 NOV 2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 30133.

XX Human: bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma

XX Homo sapiens.

XX W020157276 A2.

XX 09 AUG 2001.

XX 30 JAN 2001; 2001W0 US00668.

XX 04 FEB 2000; 2000US 0180312.

XX 26 MAY 2000; 2000US 0207456.

XX 30 JUN 2000; 2000US 0608408.

XX 03 AUG 2000; 2000US 0642366.

XX 21 SEP 2000; 2000US 0234687.

XX 27 SEP 2000; 2000US 0236359.

XX 04 OCT 2000; 2000GB 0024263.

XX (MOLE) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DG;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 30133; 658pp - Sequence listing; English.

XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.

XX Sequence 110 AA;

Query Match 3.6%; Score 7; DH 22; Length 110;

Best local similarity 100.0%; Prod. No. 29;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 LALAEKL 146

DB 82 LALA-kl 88
|||||

Search completed: July 2, 2002, 16:06:05
Job time: 142 sec

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at http://wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 110 AA;

Query Match 3.6%; Score 7; DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 LALAEKL 146
 Db 82 lalael 88
 |||||

RESULT 13
 ABB22005
 ID ABB22005 standard; Protein: 110 AA.
 XX
 AC ABB22005;
 XX
 DT 23 JAN-2002 (first entry)
 DE Protein #4004 encoded by probe for measuring heart cell gene expression.
 XX
 DE human: gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN W0200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WC-US000667.
 XX
 PP 04-FEB-2000; 2000US-0180312.
 PP 26-MAY-2000; 2000US-0207456.
 PP 30-JUN-2000; 2000US-0608408.
 PP 03-AUG-2000; 2000US-0632366.
 PP 21 SEP 2000; 2000US-0234687.
 PP 27 SEP 2000; 2000US-0236359.
 PP 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Kank DR;
 XX
 DP WPI: 2001-488899/53
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX
 PS Claim 15, SEQ ID No 23775, 530pp, English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC AHA21535-AHA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 110 AA;

Query Match 3.6%; Score 7; DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 LALAEKL 146
 Db 82 lalael 88
 |||||

RESULT 14
 AAM57432
 ID AAM57432 standard; Protein: 110 AA.
 XX
 AC AAM57432;
 XX
 DT 05-NOV-2001 (first entry)
 DE Human brain expressed single exon probe encoded protein. SEQ ID NO: 29537.
 XX
 DE Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX
 OS Homo sapiens.
 XX
 PN W0200157275-A2.
 XX
 PD 09 AUG 2001.
 XX
 PF 30-JAN-2001; 2001WC-US000667.
 XX
 PP 04-FEB-2000; 2000US-0180312.
 PP 26-MAY-2000; 2000US-0207456.
 PP 30-JUN-2000; 2000US-0608408.
 PP 04-AUG-2000; 2000US-0632366.
 PP 21-SEP-2000; 2000US-0234687.
 PP 27-SEP-2000; 2000US-0236359.
 PP 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Kank DR;
 XX
 DP WPI: 2001 483446/52
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 PS Example 4, SEQ ID No. 29537, 650pp, Sequence listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at http://wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 110 AA;

Query Match 3.6%; Score 7; DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 LALAEKL 146


```

PR 24 MAR 1992: 92GH-0006372.
PR 15 MAY 1992: 92WC-CR00884.
XX (CAMP) CAMBRIDGE ANTIBODY TECHNOLOGY.
PA (MED-) MEDICAL RES. COUNCIL.
XX
PI Baler M., Hooenboom HRJM., Jaspers USAF., Winter GP.
XX WFI: 1993-1175/14.
DR
XX producing human antibody polypeptide dimer specific for antigen .
PT comprises use of chain shuffling using phage expression, useful
PI for reducing anti globulin responses in humans for increased
PI human characteristics
XX
XX Example: Fig 11: 109pp; English.
PS
XX the sequence is that of the p0-74 VHL gene heavy chain
CC which may be used as part of a method of producing chimeric
CC mouse-human antibodies or fragments which have the same
CC binding specificity as a parent Ab but have increased human
CC characteristics, preventing anti globulin response in humans.
XX
SQ Sequence 89 AA;

```

Query Match 3.6%; Score 7; DB 14; Length 89;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 87 KPERVT 93
DB 63 KPERVT 69

```

RESULT 11
 ABB31456
 ID ABB31456 standard; Peptide; 110 AA.
 AC ABB31456;
 XX
 DT 01-FEB-2002 (first entry)
 DE Peptide #4107 encoded by breast cell single exon nucleic acid probe.
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer.
 XX
 OS Homo sapiens.
 XX
 PN W0200157271 A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001W0-0500662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236354.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001 49693/54.
 DR
 XX New spatially addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PI breast, comprises number of single exon nucleic acid probes .
 PT

Claim 27; SEQ ID NO 29303; 639pp ; sequence listing, English.

```

XX
PS
XX Claim 27; SEQ ID NO 14424; 327pp ; sequence listing, English.
CC
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and HT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. the microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: the sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIFO at http://wipo.int/pub/published\_pat\_sequences.
XX
SQ Sequence 110 AA;

```

Query Match 3.6%; Score 7; DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 140 LALAELK 146
DB 82 lalaekl 88

```

RESULT 12
 ABB36668
 ID ABB36668 standard; Peptide; 110 AA.
 AC ABB36668;
 XX
 DT 04-FEB-2002 (first entry)
 DE Peptide #4174 encoded by human foetal liver single exon probe.
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 KW Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN W0200157277 A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001W0-0500669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-483447/52.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver
 PT
 XX Claim 27; SEQ ID NO 29303; 639pp ; sequence listing, English.

```

XX AC AAM34177;
XX AC
XX DT 17-OCT-2001 (first entry)
XX DT
XX DE Peptide #8214 encoded by probe for measuring placental gene expression.
XX DE
XX KW Probe: microarray; human, placenta; antenatal diagnosis;
XX KW genetic disorder.
XX KW
XX OS Homo sapiens.
XX OS
XX PN WO200157272-A2.
XX PN
XX PD 09-AUG-2001.
XX PD
XX PF 30-JAN-2001; 2001WO-0500663.
XX PF
XX PR 04-FEB-2000; 2000US-0180312.
XX PR
XX PR 26-MAY-2000; 2000US-0207456.
XX PR
XX PR 30-JUN-2000; 2000US-0608498.
XX PR
XX PR 03-AUG-2000; 2000US-0632366.
XX PR
XX PR 21-SEP-2000; 2000US-0234687.
XX PR
XX PR 27-SEP-2000; 2000US-0236359.
XX PR
XX PR 04-OCT-2000; 2000US-0024263.
XX PR
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX PR WPI: 2001-488897/53.
XX PR
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta.
XX PT
XX PS Claim 27; SEQ ID NO 34446; 654bp, English.
XX PS
XX CC the present invention relates to single exon nucleic acid probes (SENPs)
XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX CC
XX SQ Sequence 36 AA;
XX SQ

Query Match 3.6%, Score 7; DB 22; Length 36;
Best Local Similarity 100.0%; Pred. No. 10; Mismatches 0; Gaps 0;
Matches 7; Conservative 0; Indels 0;

QY 83 GSEKFEQ 89
DB 4 ggeckfq 10
|||||

RESULT 9
AA001170
ID AA001170 standard; Protein: 83 AA.
XX AC
XX AC AA001170;
XX AC
XX DT 06-NOV-2001 (first entry)
XX DT
XX DE Human polypeptide SEQ ID NO 15062.
XX DE
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX KW
XX OS Homo sapiens
XX OS

PN WO200164835-A2.
XX PN
XX PD 07-SEP-2001.
XX PD
XX PF 26-FEB-2001; 2001WO-0504927.
XX PF
XX PR 28-FEB-2000; 2000US-0515126.
XX PR
XX PR 18-MAY-2000; 2000US-0577409.
XX PR
XX PA (HYSE-) HYSEQ INC.
XX PA
XX PI Tang YT, Liu C, Drmanac RT;
XX PI
XX PR WPI: 2001-514838/56.
XX PR
XX PR N-PSDB; AAI81101.
XX PR
XX PT Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders.
XX PT
XX FS Claim 20; SEQ ID NO 15062; 1499pp + Sequence Listing, English.
XX FS
XX CC the invention relates to human polynucleotides (AAI79941-AAI93841) and
XX CC the encoded proteins (AAU00010-AAU13910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation of which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and/or
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 83 AA;
XX SQ

Query Match 3.6%, Score 7; DB 22; Length 83;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 NYQILLK 123
DB 33 nyqillk 39
|||||

RESULT 10
AAR34271
ID AAR34271 standard; Protein: 89 AA.
XX AC
XX AC AAR34271;
XX AC
XX DT 26-JUL-1993 (first entry)
XX DT
XX DE Dp-74 VII gene heavy chain.
XX DE
XX KW Human; monoclonal antibody; anti-globulin response;
XX KW chimeric; mouse-human antibodies; antibody; prevention.
XX KW
XX OS Homo sapiens.
XX OS
XX PR W09306213-A.
XX PR
XX PD 01-APR-1993.
XX PD
XX PF 23-SEP-1992; 92WO-GR01755.
XX PF
XX PR 23-SEP-1991; 91GB 0020252.
XX PR
XX PR 25-SEP-1991; 91GB-0020377.
XX PR
XX PR 24-MAR-1992; 92GB-0006318.
XX PR

```

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: the sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from Wipo at http://wipo.int/pub/published_pat_sequences.
 XX
 XX Sequence 36 AA:

Query Match 3.6%; Score 7; DB 22; Length 36;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 GGEEKFQ 89
 DB 4 qgeekfq 10
 |||||

RESULT 6

AAM61279
 ID AAM61279 standard; Protein: 36 AA.

XX
 AC AAM61279;

XX
 DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 33384.

XX Human brain expressed exon: gene expression analysis: probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer.

XX Homo sapiens.

XX W0200157276-A2

XX 09-AUG-2001.

XX 30 JAN 2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632365.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0246359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DK;

XX WPI: 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
 CC brains -

XX Example 4: SEQ ID NO: 33384; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.

XX Sequence 36 AA:

Query Match 3.6%; Score 7; DB 22; Length 36;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 GGEEKFQ 89
 DB 4 qgeekfq 10
 |||||

RESULT 7

AAM74002
 ID AAM74002 standard; Protein: 36 AA.

XX
 AC AAM74002;

XX
 DT 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 34308.
 XX Human bone marrow expressed exon: gene expression analysis: probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX W0200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632365.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0246359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DK;

XX WPI: 2001-488900/53.

XX Human genome derived single exon nucleic acid probes useful for
 CC analyzing gene expression in human bone marrow -

XX Example 4: SEQ ID NO: 34308; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.

XX Sequence 36 AA:

Query Match 3.6%; Score 7; DB 22; Length 36;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 GGEEKFQ 89
 DB 4 qgeekfq 10
 |||||

RESULT 8

AAM34177
 ID AAM34177 standard; Protein: 36 AA.

PF 12-DEC-1997; 97JP-0343316.
XX
PR 12-DEC-1997; 97JP-0343316.
XX
PA (MARI-) MARINE BIOTECHNOLOGY INST CO LTD.
PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX
DR WPI: 1999-422615/36.
DR N-PSDB; AAX86000.
XX
PT Identification and detection of a microbe by detection of a
PT gyrase gene
XX
PS Example 3; Page 14-15; 42pp; Japanese.
XX
CC the specification describes a method for the identification or
CC detection of a microbe, using the gyrase gene as the index. The
CC method involves the use of PCR primers to amplify DNA from the
CC microbe, which is then identified or detected depending on its
CC base sequence. The method can be used to classify and identify
CC an unidentified microbe strain rapidly and with high precision.
CC The present sequence represents a gyrase protein.
XX
SQ Sequence 220 AA;
Query Match 4.1%; Score 8; DB 20; Length 220;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 SENDPEKC 39
Db 130 sendpeke 137
PFSULT 5
ABBA0457
ID ABB40457 standard; Peptide; 36 AA.
XX
AC ABB40457;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #7963 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
EF 30-JAN-2001; 2001WO-0500669
XX
PP 04-FEB-2000; 2000MS-0180312.
PP 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0637466.
PR 21-SEP-2000; 2000MS-0234687.
PR 27-SEP-2000; 2000US-0236359.
PP 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLEF-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID No 35092; 639pp + sequence listing; English.
PS

FN WO9812327-A2.
XX
PD 26-MAR-1998.
XX
PF 19-SEP-1997; 97WO-US16842.
XX
PR 04-APR-1997; 97US-0042985.
PR 20-SEP-1996; 96US-0025296.
PR 03-APR-1997; 97US-0042611.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Baer R, Bowcock AM;
XX
DR WPI: 1998 230317/20
DR N-PSDB; AAV24135
XX
XX DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which
PT as breast cancer antigen, BRCA1, binding proteins are useful to
PT identify patient having or at risk of developing cancer
XX
PS Disclosure; Page 287-288; 348pp; English.
XX
CC The sequence is that of a protein which can be used in the
CC preparation of the recombinant breast cancer antigen, BRCA1, binding
CC proteins BARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the
CC detection of a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid
CC sequence, specifically a wild type BARD1 composition for the detection
CC or purification of BRCA1, useful to identify a patient having, or at
CC risk of developing cancer. BARD1 can be used in the preparation of an
CC anti-BARD1 antibody, and in the detection and purification of a BRCA1
CC protein. BARD1, B123, BE2, BE14, BE31 or BE445 can be used in the
CC identification of a binding protein agonist or antagonist that alters
CC the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or the
CC biological activity of the BRCA1-BARD1, B123, BE2, BE14, BE31 or BE445
CC complex. The antibodies can be used to detect BARD1, B123, BE2, BE14,
CC BE31 or BE445, a specific anti-BARD1 antibody can be used to identify
CC a patient having or at risk of developing cancer.
XX
SQ Sequence 515 AA;
Query Match 48.2%; Score 94; DB 19; Length 515;
Best Local Similarity 100.0%; Pred. No. 24-89;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 102 QFSVAMADSLWKPIYQIILKTRDSSPKVRFAALITVLALAEKIKENYIVILPESIPPL 161
Db 422 qlsvamaddslwkplnyqillktrdsspkvrfaalilvlalaeikenyivilpesipl 481
QY 162 APIIMDECFERFVEHOCQRTIQQIITVLGHKPLQSYF 195
Db 482 acImdecevevheqcktiqqilvlgsp/qsyf 515
RESULT 4
AAY23793
ID AAY23793 standard; Protein; 220 AA
XX
AC AAY23793;
XX
DT 14-SEP-1999 (first entry)
XX
DE A gyrase protein of Chitinophaga pinensis.
XX
KW Identification, detection; microbe; gyrase gene; gyrase protein
XX
OS Chitinophaga pinensis.
XX
PN JPI1169175-A.
XX
PD 29-JUN-1996
XX

```

XX PI Barry C, Bouquellet L, Chumakov I, Cohen-Akenine A;
XX OR WPI: 2001-367032/38.
XX DB N-DBDB: AAF83409, AAF83410
XX New BAP28 polynucleotides and polypeptides overexpressed in prostate
XX PI cancer cells for diagnosing prostate tumors, e.g. by hybridization or
XX PI polymerase chain reaction assays -
XX Claim 14: Page 247-304; 349pp; English.
XX The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
XX sequences and regulatory region located at the 3' and 5' ends of the
XX BAP28 coding region. The BAP28 polypeptides can be expressed by standard
XX recombinant methodology. BAP28 polynucleotides and polypeptides have been
XX found to be over expressed in prostate tumor cells, therefore levels of
XX BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
XX reaction (PCR)) to diagnose patient suffering from or susceptible to
XX prostate cancer. Antibodies specific for the BAP28 polypeptides are
XX useful as diagnostic reagents. Allelic markers of the BAP28 gene are
XX useful in genetic analysis. The present sequence represents a protein
XX encoded by a first cDNA sequence of the BAP28 gene consisting of the
XX exons 1 to 45.
XX Sequence 2144 AA:
SQ
Query Match 99.0%; Score 193; DB 22; Length 2144;
Best Local Similarity 100.0%; Pred. No. 4.9e-192;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPTLAGHLVKEPAPDULXOVNISKTHFAFDSNDPEKCCLLLOPILNCILYKILPLPTQH 60
DB 1950 LPTLAGHLVKEPAPDULXOVNISKTHFAFDSNDPEKCCLLLOPILNCILYKILPLPTQH 60
QY 61 TSKPRAXALMMPLVQVQIENPLGGEEKFGEPVTKHLIPICIAQFSVAMADDSLWPLNYQI 120
DB 2010 TSKPRAXALMMPLVQVQIENPLGGEEKFGEPVTKHLIPICIAQFSVAMADDSLWPLNYQI 120
QY 121 LKTRDSSPKVRFPAALIVLALAEKIKENYIVLLPESIPPLAEIMDECHVEHQCOKTI 180
DB 2070 LKTRDSSPKVRFPAALIVLALAEKIKENYIVLLPESIPPLAEIMDECHVEHQCOKTI 180
QY 181 QQFETVLCPPLQSYF 195
DB 2130 QQFETVLCPPLQSYF 195
Query Match 65.1%; Score 122; DB 22; Length 549;
Best Local Similarity 100.0%; Pred. No. 5.2e-124;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 ALMMPLVQLENKRGSEKFGEPVTKHLIPICIAQFSVAMADDSLWPLNYQIILKTRDSS 128
DB 223 ALMMPLVQLENKRGSEKFGEPVTKHLIPICIAQFSVAMADDSLWPLNYQIILKTRDSS 128
QY 129 PKVRFAALITVLALAEKIKENYIVLLPESIPPLAEIMDECHVEHQCOKTIQOLEIVLG 188
DB 283 PKVRFAALITVLALAEKIKENYIVLLPESIPPLAEIMDECHVEHQCOKTIQOLEIVLG 188
QY 189 EPLQSYF 195
DB 343 EPLQSYF 349
RESULT 3
AAW54099
ID AAW54099 standard; Protein; 515 AA.
XX AC AAW54099;
XX XX 28-SEP-1998 (first entry)
XX DT DT
XX DE Homo sapiens BAP28 sequence.
XX KW BARD1, ring protein, BRCA1, breast cancer; risk; diagnosis.
XX OS Homo sapiens.
XX

```

GenCure version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:06:04 : Search time 117.59 seconds
(without alignments)

184,194 Million cell updates/sec

Title: US-09-603-665-5_copy_1950_2144

Perfect score: 195

Sequence: 1 LFTLFAGHLVFPFADTLXV...CKTIQLLETVLGEPLQSYF 195

Scoring table:

Gap: 60 0 0 Gapext 60 0

Searched: 747574 seqs, 11107496 residues

Word size: 6

Total number of hits satisfying chosen parameters: 699

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database:

A_Geneseq_032802:*

1:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22:	/SIDSL/qcdatala/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	99.0	2144	22	AAH85029
2	127	65.1	349	22	AAH92725
3	94	48.2	515	19	AAW54099
4	8	4.1	220	20	AAV33793
5	7	3.6	36	22	ABR40457
6	7	3.6	36	22	AAH61379
7	7	3.6	36	22	AAW74002
8	7	3.6	36	22	AAH34177
9	7	3.6	36	22	AAW01170
10	7	3.6	36	22	AAH34271
11	7	3.6	110	22	ABH31456

12	7	3.6	110	22	ABH36668
13	7	3.6	110	22	ABH22705
14	7	3.6	110	22	AAW57432
15	7	3.6	110	22	AAW69827
16	7	3.6	110	22	AAH17651
17	7	3.6	110	22	AAH30169
18	7	3.6	110	22	AAW05309
19	7	3.6	117	16	AAH66345
20	7	3.6	128	22	AAU30304
21	7	3.6	178	19	AAW37995
22	7	3.6	178	20	AAV39876
23	7	3.6	199	22	AAU39406
24	7	3.6	200	21	AAH34112
25	7	3.6	276	18	AAW55432
26	7	3.6	277	18	AAW55442
27	7	3.6	342	21	AAH42597
28	7	3.6	442	22	AAH66797
29	7	3.6	353	21	AAH43736
30	7	3.6	353	22	AAH37394
31	7	3.6	355	20	AAV36267
32	7	3.6	367	21	AAV66633
33	7	3.6	367	22	AAH65156
34	7	3.6	374	22	AAH65309
35	7	3.6	395	19	AAW57571
36	7	3.6	414	22	AAH39238
37	7	3.6	427	22	AAH41024
38	7	3.6	437	20	AAV37663
39	7	3.6	545	22	AAH81542
40	7	3.6	655	22	AAH13001
41	7	3.6	729	19	AAH47538
42	7	3.6	1045	15	AAH58611
43	7	3.6	1122	16	AAH4427
44	7	3.6	1138	21	AAH81512
45	7	3.6	1216	22	AAU7719

ALIGNMENTS

RESULT 1

ID AAB85029 standard; Protein; 2144 AA.

XX AAB85029;

UI 06-AUG-2001 (first entry)

XX Protein encoded by BAP28 cDNA consisting of exons 1 to 45.

XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1694 /label= Ser or Asn

FT Misc-difference 1854 /label= Ala or Val

FT Misc-difference 1967 /label= Asp or Asn

FT Misc-difference 2017 /label= Gly or Glu

XX WO200100669-A2.

XX 04-JAN-2001.

XX 23 JUN 2000, 2000WO-IB01183

XX 25-JUN-1999, 99US 0141323.

XX 18-JAN-2000, 2000US-0176880.

XX (GEST) GENSET.

Peptide #4174 enco
Protein #4504 enco
Human brain expres
Human bone marrow
Peptide #4385 enco
Peptide #4506 enco
Peptide #3991 enco
Human immunoglobul
Novel human secret
Mutant Aspergillus
A. oryzae 97-14.1
Novel human secret
zebra mays protein 1
H. pylori ORF hpf6
H. pylori ORF 124c
Human ORF ORF2461
Human RNA uncoilin
Human cancer assoc
Human colon cancer
Human secreted pro
Membrane-bound pro
Human PRO189 (ONQ1
Drosophila melanog
Homo sapiens BAP15
Human polypeptide
Human polypeptide
Amino acid sequenc
Human APP protein
Rice poly (A) bind
Homo sapiens BRCA1
Yeast HMG-CoA redu
Cytadhesin protein
Streptococcus pneu
Streptococcus pneu

RX MEDLINE-20083487; PubMed-10617197;
RA Lin X., Kaul S., Ponsley S.H., Shea T.P., Benito M.-L., Town C.D.,
Fuji C.Y., Mason T.M., Bowman C.L., Bonstead M.E., Fiedlikum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ponnang C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Mayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrozza A.J., Cressy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss U., Nieman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RI "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR 2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lam H., Southwick A., Karlin-Namann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Huh J., Chen H., Cheuk R.,
RA Chung M.K., Kim C., Lin J., Liu S.X., Pham P.K., Sakano H., Shinn P.,
RA Yamada K., Ecker J., Theologis A., Davis P.W.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC004332; AAB80575.1;
DR FMBI: AF370600; AAK43919.1;
DR InterPro: IPR000719; tok_pkinase.
DR Pfam: PF00669; pkinase.1
DR PROSITE: PS0011; PROTEIN_KINASE_LAM.1.
DR ATP-binding; Hypothetical protein; Transferase.
KW SEQUENCE 664 AA, 72569 MW, 4166.148;A9766 CRC64,
SQ

Query Match 4.0%; Score 8; DB 10; Length 664;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 62 KKKEGPEE 69
Db 567 KKKEGEEE 574
|||||

RESULT 14
Q97TK7 PRELIMINARY; PRT; 673 AA.
AC Q97TK7
DI 01-OCT-2001 (TrEMBLrel. 18, Created)
DI 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DI 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE "NPA, TRANSPOSASE (3' SEGMENT)".
GN CAP0093.
OS Clostridium acetobutylicum.
OG Firmicutes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE-21359325; PubMed-11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Xeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qin D., Hitti J., Wolf Y.Y.,
RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
PI "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum";
ET J. Bacteriol. 183:4823-4838(2001).
RL FMBI: AF001438; AAK76839.1;
DR InterPro: IPR002513; Transposase_7.
DR Pfam: PF01526; Transposase_7.1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 673 AA, 76570 MW, C807115C1180D07E CRC64;

Query Match 4.0%; Score 8; DB 16; Length 673;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 LDLENNKL 34
Db 207 LDLENNKL 214
|||||
RESULT 15
Q00302 PRELIMINARY; PRT; 884 AA.
AC Q00302
DI 01-JUL-1997 (TrEMBLrel. 04, Created)
DI 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DI 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ERPROT 213-21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20256781; PubMed-10794731;
RA Laplante J.M., O'Rourke P., Lu X., Pein A., Olsen A., Feinstein M.H.;
RT "Cloning of human C32: homocostasis endoplasmic reticulum protein
PT (CHERP): regulated expression of antisense cDNA depletes CHERP,
ET inhibits intracellular Ca2+ mobilization and decreases cell
ET proliferation";
RI Biochem. J. 348:189-199(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Laplante J., O'Rourke P.A., Lu X., Matthews E., Olsen A., Choi J.S.,
RA Rose E., Feinstein M.H.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U94836; AAB53327.1;
DR InterPro: IPR000467; G_patch.
DR InterPro: IPR000061; Surp.
DR Pfam: PF01585; G_patch; 1.
DR Pfam: PF01805; Surp; 1.
DR SMART: SM00443; G_patch; 1.
SQ SEQUENCE 884 AA, 100015 MW, 5188FF371HA132D9 CRC64;

Query Match 4.0%; Score 8; DB 4; Length 884;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 168 LAAIQKVV 175
Db 202 LAAIQKVV 209
|||||

Search completed: July 2, 2002, 16:18:28
Job time: 860 sec

RESULT 10
 Q94942
 ID Q94942 PRELIMINARY: PRT: 456 AA.
 AC Q94942
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DE PROBABLE UDP-N-ACETYLGLUCOSAMINE PYROPHOSPHORYLASE PROTEIN (EC 2.7.7.23).
 OS Rhizobium meliloti (S. norhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 NCBI_TaxID:482;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:1021;
 KX MEDLINE:2136824; PubMed:11474104;
 RA Galibert F., Finan T.M., Long S.R., Puchler A., Abola P., Ampe F.,
 RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
 RA Boutry M., Bowser L., Bahrmeier J., Gadieu E., Capela D., Chain P.,
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
 RA Gloux S., Gouffon E., Goffeau A., Golding B., Gouzy J., Gurjal M.,
 RA Hernandez-Lucas J., Hong A., Hualar L., Hyman R.W., Jones J., Kahn D.,
 RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Laurence V.,
 RA Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
 RA Ramsperger G., Surzycki R., Thebaud P., Vandenbol M.,
 RA Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.,
 RA "The composite genome of the legume symbiont Sinorhizobium meliloti";
 RL Science 293:668-672(2001).
 DR EMBL: AL591788; CAC46246.1;
 DR Transposase; Nucleotidyltransferase; completed; protein-
 SQ SEQUENCE 456 AA: 47775 MW: 784704BAC9C641 CRC64;

 Query Match 4.0%; Score 8; DH 15; Length 456;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 24 KKAIDLLN 41
 DB 181 KKAIDLLN 188

 RESULT 11
 Q99492
 ID Q99492 PRELIMINARY: PRT: 469 AA.
 AC Q99492
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
 DE DAN26 PROTEIN (FRAGMENT).
 CN DAN26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 KX MEDLINE:97051922; PubMed:8894557;
 RA Imort G., Sauton F., Yver G., Devys D., Trotter Y., Garnier J.M.,
 RA Weber C., Mandel J.L., Cancel G., Abbas N., Duerr A., Didierjean O.,
 RA Stevanin G., Aqid Y., Brice A.,
 RT "Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with
 RT high sensitivity to expanded CAG/glutamine repeats";
 RL Nat. Genet. 14:285-291(1996)
 DR EMBL: Y08265; CAA69591.1;
 DR InterPro: IPR000061; Surp.
 DR Pfam: PF01805; Surp: 1;
 FT NON_TER 1 1
 SQ SEQUENCE 469 AA: 52531 MW: 690F29873AFU5HIE CRC64;

Query Match 4.0%; Score 8; DH 4; Length 469;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 168 LAALQKVV 175
 DB 244 LAALQKVV 251

 RESULT 12
 Q91Q12
 ID Q91Q12 PRELIMINARY: PRT: 606 AA.
 AC Q91Q12;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DE F5D14.23 PROTEIN.
 CN F5D14.23.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsid.
 NCBI_TaxID:3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:CV. COLUMBIA;
 RA Liu S.X., Chan A., Yu G., Lee J.M., Lenz C., Pham P., Sakano H.,
 RA Iotsumi M., Vysotskaia V.S., Chin C., Chou J., Choi E., Chou M.,
 RA Pachter A., Hwang B., Liu A., Vaysberg M., Altati H., Brooks S.,
 RA Buchler E., Chao Q., Gunn S., Conway A.B., Hansen N.F., Nguyen M.,
 RA Johnson-Hopson C., Khan S., Kim C., Lam H., Miranda M., Nguyen M.,
 RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
 RA Federspiel N.A., Theologis A.;
 RT "The sequence of BAC F5D14 from Arabidopsis thaliana chromosome 1";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN:CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: AC007767; AAP81343.1;
 DR InterPro: IPR000109; PTR2.
 DR Pfam: PF00854; PTR2: 1;
 SQ SEQUENCE 606 AA: 67761 MW: 3EBCD014B19FC2F CRC64;

 Query Match 4.0%; Score 8; DH 10; Length 606;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 62 KKKEGEEE 69
 DB 5 KKKEGEEE 12

 RESULT 13
 O22808
 ID O22808 PRELIMINARY: PRT: 664 AA.
 AC O22808;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DE HYPOTHETICAL 72.6 KDA PROTEIN.
 CN ATG33580.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsid.
 NCBI_TaxID:3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:CV. COLUMBIA;

01 MAR 2001 (TRENBLREL. 16, Last annotation update)
 DE HYPOTHETICAL 12.7 KDA PROTEIN.

GN Delinococcus radiodurans.
 OS Bacteria; Thermococcus group; Deinococcus; Deinococcus.
 OX NBL_TaxID=1299;
 RN [1]

RP SEQUENCE FROM N.A.
 RS STRAIN=RL;

FX MEDLINE=20036896; PubMed=10567256;
 RA White O., Eisen J.A., Heideberg J.P., Hickey E.K., Peterson J.D.,

KA Dodson R.J., Hall D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
 RA Mollat K.S., Qin H., Chang L., Philippe W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Unterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1".

RL Sequence 286:1571-1577(1999).

DR EMBL: AE001883; AAF09F01.1; -

DR 1108; D0207; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 123 AA: 12733 MW: 3B93E87030B40810 CRC64;

Query Match 4.0% Score 8; DB 16; Length 123;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 LLSALAAL 171
 DB 5 LLSALAAL 12
 |||||

RESULT 4
 057720

ID 057720 PRELIMINARY; PRT: 123 AA.

AC 057720;

DI 01-AUG-1998 (TRENBLREL. 07, Created)

DT 01-AUG-1998 (TRENBLREL. 07, Last sequence update)

DE 01-JUN-2000 (TRENBLREL. 14, Last annotation update)

DE HYPOTHETICAL 14.2 KDA PROTEIN PH1983.

GN PH1983.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBL_TaxID=53953;

RN [1]

RP SEQUENCE FROM N.A.

RS STRAIN=OT4;

FX MEDLINE=98344137; PubMed=9679194;

KA Kawarabayashi Y., Sawada M., Horikawa H., Hatakeyama Y., Hino Y.,

KA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hesoyama A., Nagai Y.,

KA Sakai M., Ogura K., Onosaka F., Nakazawa H., Takamiya M., Ohnuki Y.,

KA Funahashi T., Tanaka T., Kaji Y., Yamazaki J., Kashiida N., Oguchi A.,

KA Aoki K. I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

KA Masuchi Y., Shizuya H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyper-

thermophilic archaeobacterium, Pyrococcus horikoshii OT3".

RL DNA Res. 5:55-76(1998).

DR EMBL: AP000007; BAA110.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 123 AA: 14212 MW: 177190F136C09A1B CRC64;

Query Match 4.0% Score 8; DB 17; Length 123;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKGGEEE 69
 DB 77 KKKGGEEE 84
 |||||

RESULT 4

ID 0950F5 PRELIMINARY; PRT: 300 AA.

AC 0950F5;

DI 01-DEC-2001 (TRENBLREL. 19, Created)

DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)

DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)

DE CYTOCHROME B (FRAGMENT).

OS Rana luteiventris.

OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.

OX NCBL_TaxID=58176;

RN [1]

RP SEQUENCE FROM N.A.

FX MEDLINE=21306528; PubMed=11412371;

RA Bos D.H., Sites J.W. Jr.;

RT "Phylogeography and conservation genetics of the Columbia Spotted Frog

(*Rana luteiventris*)".

RL Mol. Ecol. 10:1499-1513(2001).

DR EMBL: AY016673; AAK54420.1; -

KW Mitochondrion.

FT NON_TER 1 1

FT NON_TER 300 300

SQ SEQUENCE 300 AA: 33574 MW: 212A26A6CAB6BBB7 CRC64;

Query Match 4.0% Score 8; DB 8; Length 300;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 LLSALAAL 171

DB 210 LLSALAAL 217

|||||

RESULT 5

Q94PH4

ID Q94PH4 PRELIMINARY; PRT: 300 AA.

AC Q94PH4;

DI 01-DEC-2001 (TRENBLREL. 19, Created)

DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)

DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)

DE CYTOCHROME B (FRAGMENT).

OS Rana luteiventris.

OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.

OX NCBL_TaxID=58176;

RN [1]

RP SEQUENCE FROM N.A.

FX MEDLINE=21306528; PubMed=11412371;

RA Bos D.H., Sites J.W. Jr.;

RT "Phylogeography and conservation genetics of the Columbia Spotted Frog

(*Rana luteiventris*)".

RL Mol. Ecol. 10:1499-1513(2001).

DR EMBL: AY016649; AAK54396.1; -

DR EMBL: AY016652; AAK54399.1; -

DR EMBL: AY016654; AAK54401.1; -

KW Mitochondrion.

FT NON_TER 1 1

FT NON_TER 300 300

SQ SEQUENCE 300 AA: 33906 MW: 56A26A6B71HF5CA3 CRC64;

Query Match 4.0% Score 8; DB 8; Length 300;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 LLSALAAL 171

DB 210 LLSALAAL 217

|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd

OM protein - protein search, using sw model

Run on July 2, 2002, 16:04:08, Search time 96.14 seconds

(without alignments)

361.493 Million cell updates/sec

Title: US-09-603-665-5_copy_1594_1794

Perfect score: 201

Sequence: 1 LITTELEFIVEICIV:NCIP...LSYVETLSQVHLKETS 201

Scoring table: Gapped 60.0, Gapext 60.0

Searched: 562222 seqs, 17294929 residues

Word size: 6

Total number of hits satisfying chosen parameters: 2567

Minimum DH seq length: 0

Maximum DH seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmr:*
- 8: sp_organellae:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioph:*
- 17: sp_archaeap:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DH	ID	Description
1	8	4.0	108	17	Q9UXU1	Q9ux1 pyrococcus
2	8	4.0	123	16	Q9RXU9	Q9rxu9 deinococcus
3	8	4.0	123	17	O57720	O57720 pyrococcus
4	8	4.0	300	8	O950F5	O950f5 rana lutely
5	8	4.0	300	8	O94PH4	O94ph4 rana lutely
6	8	4.0	307	4	Q9AAG4	Q9aag4 homo sapien
7	8	4.0	343	2	Q9K3Y1	Q9k3y1 streptomyce
8	8	4.0	394	3	Q96TY7	Q96ty7 neurospora
9	8	4.0	439	11	Q91W08	Q91w08 mus musculu
10	8	4.0	456	16	O92PS3	O92ps3 rhizobium m
11	8	4.0	468	4	Q944G2	Q944g2 homo sapien
12	8	4.0	606	10	Q9LQJ2	Q9lqj2 arabidopsis
13	8	4.0	664	10	O2280W	O2280w arabidopsis
14	8	4.0	673	16	Q97TK7	Q97tk7 clostridium
15	8	4.0	884	4	O00302	O00302 homo sapien
16	8	4.0	9507	2	Q9EWA1	Q9ewa1 streptomyce

ALIGNMENTS

RESULT 1

Q9UXU1

IN Q9UXU1 PRELIMINARY; PPT; 108 AA

AC Q9UXU1

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE HYPOHETICAL 12.5 KDA PROTEIN.

CN PAB179.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI_TaxID:29292;

RN [1]

RP SOURCE FROM N.A.

RC STRAIN=ORSAY;

RA Heilig K.;

FT structure and evolution.*;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ248288; CAH50672.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 108 AA; 12527 MW; 170274A4A8A659C1 CRC64;

Query Match 4.0%; Score 8, DB 17, Length 108.

Best local Similarity 100.0%; Pred. No. 5.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKEEEE 69

DB 62 KKKKEEEE 69

RESULT 2

Q9RXU9

ID Q9RXU9 PRELIMINARY; PPT; 123 AA.

AC Q9RXU9

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```

FT DOMAIN 7 42 POLY-GLU.
FT NON-TER 420 420
SQ SEQUENCE 420 AA: 420 MW: 114048 [HE86A0EF0 CRC64;]

Query Match 3.5% Score 7; DB 1; Length 454;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 VISTLEA 144
DB 402 VISTLEA 408

RESULT 15
CHAZ SCHHO STANDARD PRI: 354 AA.
AP 04665;
DT 01 JUN 1994 (Rel. 29, Last sequence update)
DT 01 JUN 1994 (Rel. 29, Last sequence update)
DT 16 OCT 2001 (Rel. 40, Last annotation update)
DE Guanine nucleotide-binding protein alpha-2 subunit (G12-alpha).
GN G12 OR SPAC2484.130.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
FN [1]
RP SEQUENCE FROM N.A. AND MUTAGENESIS OF AP3-176 AND GLN-202.
EX MEDLINE-94040710; PubMed-140462;
RA IISHIKI T., MOCHIZUKI N., MAEDA T., YAMAMOTO M.;
RT "Characterization of a fission yeast gene, gpa2, that encodes a G
RT alpha subunit involved in the monitoring of nutrition.";
RL Genes Dev. 6:2455-2462(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Skolton J., Churcher C.M., Barrell H.G., Rajandream M.A., Walsh S.V.,
RA Wood V.;
RL Submitted (SEP 1997) to the EMBL/GenBank/DDBJ databases.
CC 1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS. THIS PROTEIN MAY BE INVOLVED IN THE
CC DETERMINATION OF THE CAMP LEVEL ACCORDING TO NUTRITIONAL
CC CONDITIONS, MOST PROBABLY AS A REGULATOR OF ADENYL CYCLASE.
CC 1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.
CC 1- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DB EMBL: U03366; XA02630.1;
DB EMBL: Z99163; CAB16244.1;
DB EMBL: A46493; A46393.
DB HSSP: F10424; IAS3.
DB InterPro: IPR001019; Gprotein_alpha.
DB Pfam: PF00503; G_alpha_1.
DB PRINTS: PR00418; GPROTEIN.
DB SMART: SM00275; G-alpha_1.
KW GTP binding; Transducer; Multigene family.
FT NP_BIND 41 48 GTP (BY SIMILARITY).
FT NP_BIND 194 202 GTP (BY SIMILARITY).
FT NP_BIND 269 272 GTP (BY SIMILARITY).
FT MUTAGEN 176 176 R->H: DECREASE IN MATING AND SPORULATION
FT EFFICIENCY.
FT MUTAGEN 202 202 Q->L: DECREASE IN MATING AND SPORULATION

```

```

FT SEQUENCE 354 AA: 40532 MW: A0042174 (138988A CRC64;]
EFFICIENCY.

Query Match 3.5% Score 7; DB 1; Length 454;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TRIPVIR 12
DB 69 TRIPVIR 75

Search completed: July 2, 2002, 16:19:05
Job time: 777 sec

```

OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RA MEDLINE 21145866; PubMed=11248103;
 RX May B.J., Zhang Q., Li L.L., Eastham M.L., Whittam T.S., Kapur V.,
 RT "Complete genomic sequence of *Enterococcus faecalis* PM70".
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -!- FUNCTION: THIS ENZYME MAY PLAY A SIGNIFICANT ROLE IN PROCESSES
 CC LEADING TO RECOVERY FROM MUTAGENESIS AND/OR CELL DEATH BY
 CC ALKYLATING AGENTS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N7-
 CC methylguanine residues, releasing 2,6-diamino 4 hydroxy 5 (N-
 CC methyl)formamidopyrimidine.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE FPG FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC EMBL: AE006155; AAC03229.1;
 DR InterPro: IPR000191; Fapy_DNA_glyco.
 DR InterPro: IPR000214; Fapy_DNAGlyco_gn.
 DR Pfam: PF01149; Fapy_DNA_glyco, 1.
 DR Proxam: PD03680; Fapy_DNA_glyco, 1.
 DR PROSITE: PS01242; FPG; 1.
 KW DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
 KW Complete proteome.
 FT ZN-FING 244 267 POTENTIAL.
 FT SPOUNCE 270 AA; 30612 MW; H4C5HDFD968734 CRC64;
 SQ
 Query Match 3.5%; Score 7; DB 1; Length 270;
 Best local Similarity 100.0%; Pred No. 26;
 Matches 7; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
 OY 182 ISPYLEG 188
 Db [1]
 14 ISPYLEG 20
 RESULT 13
 PH86 YEAST
 ID PH86 YEAST STANDARD; PRT; 311 AA.
 AC P46956;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE Inorganic phosphate transporter PH086.
 GN PH086 OR YJL117W OR J0744.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRF88;
 RX MEDLINE=96171517; PubMed=8598055;
 RA Hun-Ya M., Shikata K., Nakade S., Vompakdee C., Harashima S.,
 KA Oshima Y.;
 RT "Two new genes, PH086 and PH087, involved in inorganic phosphate
 RT uptake in *Saccharomyces cerevisiae*.";
 RL Curr. Genet. 29:344-351(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FV1679;
 RX MEDLINE=87103775; PubMed=8948101.
 RA Czepluch C., Kordes E., Pujol A., Taniaux J.-C.;

RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
 RT reveals 19 open reading frames including URA2 (5' end), TRP1, PBS2,
 RT SPI10, CCB14, RPL1, PH086, NCA3, ASF1, CCB7, GZF3, two tRNA genes,
 RT three remnant delta elements and a Ty4 transposon.";
 CC Yeast 12:1471-1474(1996)
 CC -!- FUNCTION: INVOLVED IN THE UPTAKE OF INORGANIC PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC EMBL: D04817; HAA09885.1;
 DR EMBL: Z4932; CAH9412.1;
 DR SCD; S0004653; PH086.
 KW Phosphate transport; Transport; Transmembrane.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 118 138 POTENTIAL.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 180 180 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SPOUNCE 311 AA; 34881 MW; ICR19C9184BFC48 CRC64;
 SQ
 Query Match 3.5%; Score 7; DB 1; Length 311;
 Best local Similarity 100.0%; Pred No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 60 ORKKKKG 66
 Db [1]
 5 ORKKKKG 11
 RESULT 14
 HAP1 HUMAN
 ID HAP1 HUMAN STANDARD; PRT; 320 AA.
 AC P54257;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE HAP1-like protein 1 (Fragment).
 GN HAP1.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Caudate;
 RX MEDLINE=96077112; PubMed=7477378;
 RA Li X.-J., Li S.-H., Sharp A.H., Nucleifera F.C. Jr., Schilling G.,
 PA Lanahan A., Worley P., Snyder S.H., Ross C.A.;
 RT "A huntingtin-associated protein enriched in brain with implications
 RT for pathology.";
 RL Nature 378:398-402(1995).
 CC -!- FUNCTION: IN CONTRAST TO HAP1, DOES NOT SEEM TO BIND TO
 CC HUNTINGTIN.
 CC -!- SIMILARITY: SJFING, TO HAP1
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U38371; AAC50297.1;
 DR NON_TER 1 1

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL: A001094; AAB01063.1; 1;
ID: F0172;
KW: Hypothetical protein; Transmembrane; Complete proteome;
FT: TRANSMEM 13 35 POTENTIAL;
FT: TRANSMEM 50 68 POTENTIAL;
FT: TRANSMEM 75 94 POTENTIAL;
FT: TRANSMEM 109 131 POTENTIAL;
FT: TRANSMEM 152 174 POTENTIAL;
FT: TRANSMEM 194 216 POTENTIAL;
SQ: SEQUENCE: 217 AA; 2-847 MW; 911480F0267C8B6 CRC64;

Query Match 3.5%; Score 7; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 LLSALAA 170
DB 57 LLSALAA 64
|||||

RESULT 10
ID YEAST STANDARD; PRT: 218 AA;
AC P40040;
DI 01-FEB-1995 (Rel. 41, Created)
DI 01-FEB-1995 (Rel. 41, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein.
GN TPO1 OR YER064W.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Euk; Ascomycota; Saccharomycetia; Saccharomycetes.
OC Saccharomycetes; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4912;

[1]
SEQUENCE FROM N.A.
PP STRAIN: S288C / AB672.
RA Dietrich F.S., Mullen J.L., Hennessey K.M., Allen R., Araujo R., Aviles E., Bono A., Brennan P., Carpenter J., Chen P., Cherry J.M., Chung K., Duncan M., Guzman E., Harizell G., Hunkeler-Smith S., Hyman R., Kayser A., Komp C., Laskari D., Lee H., Lin D., Maseda D., Nakahara K., Namah A., Norgran P., Ojner P., Oh G., Patel F.X., Roberts D., Sehl P., Sehrman S., Shogren T., Smith V., Taylor P., Wei Y., Yelton M., Hefstein D., Davis R.W.;
RL Submitted (08-1994) to the EMBL/Genbank/DDBJ databases [2]

CHARACTERIZATION:
RP MEDLINE=9483750; PubMed=9707445;
RA Pirnat J.L., Aquilera A.;

RA "A novel yeast gene, TPO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination."

RL EMBL J. 17:4859-4874(1998).
OC FUNCTION: UNKNOWN; SUPPRESSOR OF THE TRANSCRIPTIONAL DEFECT OF HFR1 BY OVEREXPRESSION.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL: J18814; AAB64599.1;
ID SGD: S000085; YER064W.

DR InterPro: IPR004034; SAP;
DR Pfam: PF02037; SAP: 1;
DR SMART: SM00514; SAP: 1;
SQ SEQUENCE: 218 AA; 24137 MW; 392B857BAFBA94A CRC64;

Query Match 3.5%; Score 7; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KALDLIN 31
DB 128 KALDLIN 134
|||||

RESULT 11
ID PHOO_ZYMMO STANDARD; PRT: 237 AA;
AC Q9X5E1;
DI 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphate transport system protein pho0.
GN PHO0.

OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN: ATCC 31821 / ZM4 / CP4;
RA Lee H.J., Kang H.S.;
RL Submitted (1AN-1994) to the EMBL/Genbank/DDBJ databases.
CC FUNCTION: NOT KNOWN; PROBABLY INVOLVED IN PHOSPHATE TRANSPORT AND/OR METABOLISM (BY SIMILARITY).
CC SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC SIMILARITY: BELONGS TO THE PHO0 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL: AF124757; AAD29648.1;
DR Phosphate transport.
SQ SEQUENCE: 237 AA; 26270 MW; 54549210E48D892B CRC64;

Query Match 3.5%; Score 7; DB 1; Length 237;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 TSTLEAL 135
DB 37 TSTLEAL 43
|||||

RESULT 12
ID FPG_PASMO STANDARD; PRT: 270 AA;
AC P57910;
DI 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE Formamido-pyrimidine DNA glycosylase (Ec 3.2.2.23) (Fap/ DNA glycosylase).
GN MUTM OR FPG OR PM1145.

OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.

CC -!- FUNCTION: ALSO STIMULATES PTS2 EXPRESSION. THIS SUGGESTS THAT RCSEB
CC IS A REGULATOR IMPLICATED IN THE REGULATION OF MORE THAN ONE
CC CELLULAR FUNCTION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE LUXR/DHXA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -!- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X87840; CAA61094.1;
CC EMBL: AB000683; BAA19163.1;
CC EMBL: A1627274; CAD07501.1;
CC InterPro: IPR0001792; HTH_LuxR
CC InterPro: IPR001789; Response_Reg.
CC Pfam: PF00196; GntE, 1.
CC Pfam: PF00072; response_reg, 1.
CC PRINTS: PP00004; HTH_LuxR.
CC SMART: SM00421; HTH_LuxR, 1.
CC SMART: SM00448; REC, 1.
CC PROSITE: PS00622; HTH_LuxR_FAMILY, 1.
CC PROSITE: PS00110; RESPONSE_REGULATORY, 1.
CC Sensory Transduction, Phosphorylation, Transcription regulation,
CC DNA-binding, Activator, Complete proteome.
CC DOMAIN 1 124 RESPONSE REGULATORY.
FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).
FT DNA_BIND 168 187 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 216 AA; 23728 MW; D35CE2AB2BDC0E4 CRC64;

Query Match 3.5% Score 7; DB 1; Length 216;
Best local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ALAALQK 173
|||||
DB 119 ALAALQK 125

RESULT 8
RCSEB_SALT
ID RCSEB_SALT STANDBY PRT; 216 AA.
AC P58663;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Capsular synthesis regulator component B.
GN RCSEB OR STM2270.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria, gamma subdivision, Enterobacteriaceae;
OC Salmonella.
CX NCBI_TaxID:602;
RN [1]
PP SEQUENCE FROM N A
PP STRAIN LT2 / SC5C1412 / ATCC 13075.
EX MEDLINE:21534948; PubMed:1577609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
PA Courtney L., Porwolik S., Ali J., Dainton M., Du F., Hou S., Layman D.,
PA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2".
PL Nature 413:852-856(2001).

CC -!- FUNCTION: RCSEB IS A MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM,
CC RCSEB/RCSC, WHICH REGULATES THE EXPRESSION OF GENES INVOLVED IN
CC COLANIC ACID CAPSULE SYNTHESIS. RCSEB ACTS AS THE EFFECTOR.
CC ADDITIONAL CONTROL IS PROVIDED BY THE DEPENDENCE ON THE ALTERNATE

CC SIGMA FACTOR, RPN, FOR THE SYNTHESIS OF RCSEB. RCSEB AND RCSEB FORM
CC A COMPLEX TO PROMOTE TRANSCRIPTION OF THE GENES FOR CAPSULE
CC SYNTHESIS.
CC -!- FUNCTION: ALSO STIMULATES PTS2 EXPRESSION. THIS SUGGESTS THAT RCSEB
CC IS A REGULATOR IMPLICATED IN THE REGULATION OF MORE THAN ONE
CC CELLULAR FUNCTION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE LUXR/DHXA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -!- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF008801; AAI21171.1;
CC ST7Gene, SG7377; RCSEB.
CC PROSITE: PS00622; HTH_LuxR_FAMILY, 1.
CC PROSITE: PS00110; RESPONSE_REGULATORY, 1.
CC Sensory Transduction, Phosphorylation, Transcription regulation,
CC DNA-binding, Activator, Complete proteome.
CC DOMAIN 1 124 RESPONSE REGULATORY.
FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).
FT DNA_BIND 168 187 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 216 AA; 23714 MW; A63D1E7F704E9589 CRC64;

Query Match 3.5% Score 7; DB 1; Length 216;
Best local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ALAALQK 173
|||||
DB 119 ALAALQK 125

RESULT 9
Y172_ARCFU
ID Y172_ARCFU STANDBY PRT; 217 AA
AC O30065;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0172.
GN AF0172.
OS Archaeoglobus fulgidus
OC Archaea; Euryarchaeota; Archaeoglobales, Archaeoglobaceae;
OC Archaeoglobus
CX NCBI_TaxID:2234;
RN [1]
PP SEQUENCE FROM N.A.
PP STRAIN VC-16 / DSM 4304 / ATCC 49558;
EX MEDLINE:98049343; PubMed:9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Klenk H.-P., Clayton R.A., Tomb J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.P., Graham D.E., Kyrpides N.C.,
RA Fleischmann P.D., Clark-Kraus J., Iyer H., Sutton G.G., Gill S.,
RA Kirkness P.F., Dougherty R.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Hutch C.L., McNeil L.K., Badger J.H., Glick A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olson G.J., Fraser C.M., Smith W.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphato-
PP reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mao B., Shao Y.,
RA "The complete genome sequence of *Escherichia coli* K-12.";
RA Science 277:1234-1238 (1997).
RA [4]
RA SEQUENCE FROM N.A.
RA STRAIN-K12;
RA MEDLINE-97251358; PubMed-9037040;
RA Itoh T., Aiba H., Baba J., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitatani M.,
RA Makino K., Miki J., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakado S., Nakamura Y., Nishimoto H., Nishio Y., Oshima H.,
RA Saito N., Sempel G., Seki Y., Silvasundaram S., Tadami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.,
RA "A 460 kb DNA sequence of the *Escherichia coli* K-12 genome
RA corresponding to the 40.1-50.0 min region on the linkage map.";
RA DNA Res. 3:379-392 (1996).
RA [5]
RA SEQUENCE FROM N.A.
RA STRAIN-0157-H7 / ECU-4 / ATCC 700927;
RA MEDLINE-21074935; PubMed-11206551;
RA Perna N.F., Plunkett G., Litt H., Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink A., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.,
RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
RA Nature 409:529-533 (2001).
RA [6]
RA SEQUENCE FROM N.A.
RA STRAIN-0157-H7 / RMD 0505952;
RA MEDLINE-21156231; PubMed-11238796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo K., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.,
RA "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
RA O157:H7 and genomic comparison with a laboratory strain K-12.";
RA DNA Res. 8:111-122 (2001).
RA [7]
RA FUNCTION: RCSB IS A MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM,
RA RCSC/RCSB, WHICH REGULATES THE EXPRESSION OF GENES INVOLVED IN
RA COLANIC ACID CAPSULE SYNTHESIS. RCSC ACTS AS THE EFFECTOR
RA ADDITIONAL CONTROL IS PROVIDED BY THE DEPENDENCE ON THE ALTERNATE
RA SIGMA FACTOR, RPOH, FOR THE SYNTHESIS OF RCSC. RCSC AND RPOH FORM
RA A COMPLEX TO PROMOTE TRANSCRIPTION OF THE GENES FOR CAPSULE
RA SYNTHESIS.
RA [8]
RA FUNCTION: ALSO STIMULATES FTSZ EXPRESSION. THIS SUGGESTS THAT RCSC
RA IS A REGULATOR IMPLICATED IN THE REGULATION OF MORE THAN ONE
RA CELLULAR FUNCTION.
RA [9]
RA SIMILARITY: BELONGS TO THE LUXR/UNRA FAMILY OF TRANSCRIPTIONAL
RA REGULATORS.
RA [10]
RA SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
RA [11]
RA This SWISS-PROT entry is copyrighted by the EMBL, which is a collaboration
RA between the Swiss Institute of Bioinformatics and the EMBL outstation
RA at the European Bioinformatics Institute. There are no restrictions on its
RA use by non-profit institutions as long as its content is in no way
RA modified and this statement is not removed. Usage by and for commercial
RA entities requires a license agreement (See <http://www.ish-sib.ch/announce/>
RA or send an email to license@sib.ch).
RA [12]
RA EMBL: M28242; AAA24504.1;
RA EMBL: L11272; AAA24506.1;
RA EMBL: AF000310; AA75227.1;
RA EMBL: D90850; BAA16000.1;
RA EMBL: D90851; BAA16308.1;
RA EMBL: AE005453; AA657352.1;
RA EMBL: AP002560; BAB36529.1;
RA PIR: J40068; BV67CH.
RA EMBL: EG10821; RCSB.
RA InterPro: IPR000792; HTH_LuxR.
RA InterPro: IPR001789; Response_reg.

DR Pfam: PF00196; GENE: 1.
DR Pfam: PF00072; response_reg: 1.
DR PRINTS: PR00048; HTH_LuxR.
DR SMART: SM00421; HTH_LuxR: 1.
DR SMART: SM00448; REC: 1.
DR PROSITE: PS00622; HTH_LuxR_FAMILY: 1.
DR PROSITE: PS0110; RESPONSE_REGULATOR: 1.
DR Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Activator; Complete proteome.
FT DOMAIN 1 124 RESPONSE REGULATORY.
FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).
FT DNA_BIND 168 187 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 216 AA; 23670 MW; A78D1RD3004E0680 CRG64;

Query Match 3.5%; Score 7; DH 1; Length 216;
Rest Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ALAALQK 173
DB 119 ALAALQK 125
IIIIII

RESULT 7
RCSB_SALT
ID RCSB_SALT STANDARD; PRI: 216 AA.
AC Q56127;
DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE Capsular synthesis regulator component B.
GN RCSB OR STV2495.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteriella.
OX NCBI_TaxID-601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TY2;
RX MEDLINE-96198173; PubMed-8626298;
RA Virloquet J., Waxin H., Ecobichon C., Lee J.O., Popoff M.Y.;
RA "Characterization of the rcsA and rcsB genes from *Salmonella typhi*;
RA rcsB through typhA is involved in regulation of Vi antigen
RA synthesis.";
RA J. Bacteriol. 178:1691-1698 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-GIF0 10007;
RA Hashimoto Y.;
RN [3]
RP Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CT18;
RX MEDLINE-21534947; PubMed-11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.K., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.H.G., Sebahia M.,
RA Baker S., Hasham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Fellwell T., Hamlin N., Hance A., Hien T.T., Bellamy S., Jaisels K.,
RA Krog A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skellon J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RA "Complete genome sequence of a multiple drug resistant *Salmonella*
RA enterica serovar Typhi CT18.";
RA Nature 413:848-852 (2001).
CC [1] FUNCTION: RCSB IS A MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM,
CC RCSC/RCSB, WHICH REGULATES THE EXPRESSION OF GENES INVOLVED IN
CC COLANIC ACID CAPSULE SYNTHESIS. RCSC ACTS AS THE EFFECTOR.
CC ADDITIONAL CONTROL IS PROVIDED BY THE DEPENDENCE ON THE ALTERNATE
CC SIGMA FACTOR, RPOH, FOR THE SYNTHESIS OF RCSC. RCSC AND RPOH FORM
CC A COMPLEX TO PROMOTE TRANSCRIPTION OF THE GENES FOR CAPSULE
CC SYNTHESIS.

Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: MAY PLAY A ROLE IN L-LACTATE TRANSPORT.
 CC - SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC - SIMILARITY: BELONGS TO THE LLDLP FAMILY OF TRANSPORTERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: Z24043, CAB08022.1;
 DR EMBL: Z71928; CA936486.1;
 DR EMBL: Z99121; CAB15424.1;
 DR Subtilist: HG11875; yvfh.
 DR InterPro: IPR003804; Lactate_perm.
 DR Pfam: PF02652; Lactate_perm; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 37 57 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 249 269 POTENTIAL.
 FT TRANSMEM 304 324 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 FT TRANSMEM 448 468 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 542 562 POTENTIAL.
 SQ SEQUENCE 563 AA: 59761 MW: 638098.05949808 CKB04;

Query Match 4.0%; Score 8; DB 1; Length 563;
 Best local Similarity 100.0%; Pred No 5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 164 LLSALAL 171
 Dd | | | | | | | |
 15 LLSALAL 22

RESULT 5
 YBAW_ECOLI
 ID YBAW_ECOLI STANDARD; PRT; 132 AA.
 AC P77712;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ybaW.
 GN YBAW OR B0443.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision, Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor T., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose J.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RT Science 277:1453-1474 (1997).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP SUPRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor T., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose J.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RT Science 277:1453-1474 (1997).
 RL [2]

Query Match 3.5%; Score 7; DB 1; Length 132;
 Best local Similarity 100.0%; Pred No 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 188 GILSQVI 194
 Dd | | | | | | | |
 84 GILSQVI 90

RESULT 6
 RGSB_ECOLI
 ID RGSB_ECOLI STANDARD; PRT; 216 AA.
 AC P14374;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAP-2002 (Rel. 41, Last annotation update)
 DE Capsular synthesis regulator component B.
 GN RGSB OR R2217 OR Z3476 OR PRC5106
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision, Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-90130299; PubMed-2404948;
 RA Strout V., Gottesman S.;
 RT "RgsB and RgsC: a two-component regulator of capsule synthesis in
 RT Escherichia coli."
 RT J. Bacteriol. 172:659-669 (1990).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STPAIN-06-K30:H12;
 RC MEDLINE-93374832; PubMed-8366025;
 RX Jayaratne P., Keenleyside W.J., MacLachlan P.R., Dodgson C.,
 RA Whitfield C.;
 RT "Characterization of rcsB and rcsC from Escherichia coli O9-K30:H12
 RT and examination of the role of the rcs regulatory system in
 RT expression of group I capsular polysaccharides."
 RT J. Bacteriol. 175:5384-5394 (1993).
 RL [3]
 RN SEQUENCE FROM N.A.
 RP SUPRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,

```

FI  VARIANI 2017 2017 /FTID-VAR_010941.
FI  VARIANI 2017 2017 /FTID-VAR_010942.
SI  SEQUENC 2144 AA: 242355 MW: D66815EE78C8B7 CRC64;

Query Match 49.8%; Score 100; DH 1; Length 2144;
Best Local Similarity 100.0%; Pred. No. 4,4e-41;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPTETPTPTVIRGLVGVPLSVRSKALDLNNKLNQNISSKKTIVTRFKLVPILLAIQ 60
DB 1 LPTETPTPTVIRGLVGVPLSVRSKALDLNNKLNQNISSKKTIVTRFKLVPILLAIQ 1654
DB 1 LPTETPTPTVIRGLVGVPLSVRSKALDLNNKLNQNISSKKTIVTRFKLVPILLAIQ 1654

QY 61 KKKKEGEEQAINQTAITYTLKLLKKNFGAENPDPFVPL 100
DB 1 KKKKEGEEQAINQTAITYTLKLLKKNFGAENPDPFVPL 1693
DB 1 KKKKEGEEQAINQTAITYTLKLLKKNFGAENPDPFVPL 1693

RESULT 2
ID BP28 MA:FA STANDARD: PRT: 958 AA.
AC Q6GM44;
DT 16 OCT-2001 (Rel. 40, Created)
DI 16 OCT-2001 (Rel. 40, Last sequence update)
DE Protein BAP28 (Fragment).
OS Macaca fascicularis (Orang-eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC ISSUES-Brain;
RA Osada N., Hida M., Kusaka T., Iizuma K., Iseki Y., Hirai M., Iwano K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT *Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.*;
RI Submitted (JUL 2000) to the EMBL/GenBank/NCBI databases.
CC -1- SIMILARITY: BELONGS TO THE BAP28 FAMILY.
CC -1- SIMILARITY: CONTAINS 1 HEAT REPEAT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: AB049842; BAB16728.1; ALT_INIT.
DR InterPro: IPR000457; HEAT_Repeat.
DR PROSITE: PS00077; HEAT_REPEAT; FALSI_NEG.
FI NON_ITER 1
FI REPEAT 920 956 HEAT.
SI SEQUENCE 958 AA: 108644 MW: 40809503624CFB31 CRC64;

Query Match 49.8%; Score 82; DH 1; Length 958;
Best Local Similarity 100.0%; Pred. No. 1,5e-74;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LPSVRRKALDINNKIQINISWKKITIVTRFKLVPILLAIQVKKKKEGEEQAINQTAITY 78
DB 425 LPSVRRKALDINNKIQINISWKKITIVTRFKLVPILLAIQVKKKKEGEEQAINQTAITY 484

QY 79 YTLKLLKKNFGAENPDPFVPL 100
DB 485 YTLKLLKKNFGAENPDPFVPL 506

```

```

RESULT 3
ID YA97 MYCPN STANDARD: PRT: 541 AA.
AC P75595;
DT 16 OCT-2001 (Rel. 40, Created)
DI 16 OCT-2001 (Rel. 40, Last sequence update)
DI 16 OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein MPN097 precursor (R02_orf541).
GN MPN097 OR MP057
OS Mycoplasma pneumoniae
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:ATCC 29342 / M129;
RC MEDLINE-97105885; PubMed-8048633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li H.-C.,
RA Herrmann R.;
RT *Complete sequence analysis of the genome of the Bacterium Mycoplasma
RT pneumoniae.*;
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC -1- SIMILARITY: BELONGS TO THE MGI85 / MC260 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: AE000007; AAB95705.1;
KW Hypothetical protein, lipoprotein; Membrane; Signal;
KW Complete proteome.
FI SIGNAL 1 22 POTENTIAL.
FI CHAIN 23 541 HYPOTHETICAL LIPOPROTEIN MPN097.
FI LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SI SEQUENCE 541 AA: 59153 MW: 4247328626B1E9C CRC64;

Query Match 4.0%; Score 8; DH 1; Length 541;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 AALQKWE 176
DB 50 AALQKWE 57

RESULT 4
ID YVPH_BACSU STANDARD: PRT: 563 AA.
AC P71067;
DT 01-NOV-1997 (Rel. 35, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DI 16 OCT-2001 (Rel. 40, Last annotation update)
DE Putative D-lactate permease YVPH.
GN YVPH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC Denizot F.C.;
RI Submitted (APR 1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 51-563 FROM N.A.
RC STRAIN-168;
RA Fabret C., Quentin Y., Chapel N., Guiseppe A., Haicoh J., Denizot F.;

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on July 2, 2002, 16:06:38 : Search time: 19 seconds
(without alignments)

257,788 Million cell updates/sec

Title: US-09-603-665-5_COPY_1594_1794

Perfect score:

Sequence:

1 LPTETETPVRIGVGNP . . . ISPVLEGLISQVTHLEKITS 201

Scoring table:

Gapop 50.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 6

Total number of hits satisfying chosen parameters: 514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	49.8	2144	1	BP28_HUMAN
2	82	40.8	958	1	Q9H593 homo sapien
3	8	4.0	541	1	Q9H593 macaca fasc
4	8	4.0	541	1	Q9H593 mycoplasma
5	7	3.5	132	1	Q9H593 bacillus su
6	7	3.5	132	1	Q9H593 escherichia
7	7	3.5	132	1	Q9H593 escherichia
8	7	3.5	132	1	Q9H593 salmonella
9	7	3.5	132	1	Q9H593 salmonella
10	7	3.5	132	1	Q9H593 archaeoglob
11	7	3.5	132	1	Q9H593 saccharomyc
12	7	3.5	132	1	Q9H593 symonoma m
13	7	3.5	132	1	Q9H593 pasteurella
14	7	3.5	132	1	Q9H593 saccharomyc
15	7	3.5	132	1	Q9H593 homo sapien
16	7	3.5	132	1	Q9H593 schizosacch
17	7	3.5	132	1	Q9H593 mycobacteri
18	7	3.5	132	1	Q9H593 streptococ
19	7	3.5	132	1	Q9H593 streptococ
20	7	3.5	132	1	Q9H593 homo sapien
21	7	3.5	132	1	Q9H593 escherichia
22	7	3.5	132	1	Q9H593 sus scrofa
23	7	3.5	132	1	Q9H593 sus scrofa
24	7	3.5	132	1	Q9H593 porphyromon
25	7	3.5	132	1	Q9H593 homo sapien
26	7	3.5	132	1	Q9H593 eimeria ten
27	7	3.5	132	1	Q9H593 mycoplasma
28	7	3.5	132	1	Q9H593 escherichia
29	7	3.5	132	1	Q9H593 mus musculus
30	7	3.5	132	1	Q9H593 bos taurus
31	7	3.5	132	1	Q9H593 homo sapien
32	6	3.0	52	1	Q9H593 trypanosoma
33	6	3.0	52	1	Q9H593 trypanosoma

34 6 3.0 55 1 RS21_HACSI
35 6 3.0 57 1 RS21_HACHD
36 6 3.0 62 1 YM45_FAEET
37 6 3.0 76 1 CONG_CONGE
38 6 3.0 77 1 TATL_AQUIAE
39 6 3.0 79 1 IPK1_HUMAN
40 6 3.0 79 1 IPK1_RAT
41 6 3.0 79 1 IPK3_RAT
42 6 3.0 80 1 ISK3_MOUSE
43 6 3.0 83 1 PAD4_ECOLI
44 6 3.0 84 1 ACP_PORPO
45 6 3.0 91 1 MERP_PSEPL

p23829 bacillus st
Q9K465 bacillus ba
p34521 caenorhabdi
Q9XV55 conus geor
Q66478 aquilex aeo
p00955 homo sapien
p09655 rattus norv
p09656 rattus norv
p09036 mus musculu
p22995 escherichia
p51280 porphyra pu
Q51770 pseudomonas

ALIGNMENTS

RESULT 1
BP28_HUMAN
ID BP28_HUMAN STANDARD; PRT; 2144 AA.
AC Q9H593; Q9NW23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein BAP28.
CN BAP28.
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini, Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT S 1691; A-1854; D-1967 AND C-2017.
RA Bouquellier L., Chumakov I., Barry C., Cohen-Akenine A.;
RT "A novel BAP28 gene and protein.";
RL Patent number W001006669, 04-JAN-2001.
RN [2]
RP SEQUENCE OF 1534-2144 FROM N.A.
RA Cobley V.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1777-2144 FROM N.A.
RA Isogai T., Ito T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
Wagatsuma M., Hosokawa Y., Kaku Y., Kudo H., Kudo H.,
Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
Ninomiya K., Iwayanagi T.;
RT "New human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO THE BAP28 FAMILY.
CC - SIMILARITY: CONTAINS 1 HEAT REPEAT.

This SWISS-PROT entry is copyright © 1993-2000 by the EMBL/GenBank/DBJ databases. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL/GenBank/DBJ databases. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/autocore/> or send an email to license@isb-sib.ch).

EMBL: AK067150; CAC26776.1;
EMBL: AL136105; CAC15948.1;
EMBL: AK001221; BAA91564.1; ALT INIT.
InterPro: IPR000357; HEAT_repeat
PROSITE: PS00077; HEAT_REPEAT; FALSE_NCG.
Polymorphism: 2106 2142 N -> S.
REPEAT 1694 1694
VARIANT 1694 1694 /FTId-VAR_010949.
VARIANT 1694 1694 V -> A.
VARIANT 1697 1697 /FTId-VAR_010940
VARIANT 1697 1697 N -> D

Best Local Similarity 100.0%; Score 7; DH 2; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 GILSVI 194
IIIIII
DB 44 GILSVI 90

RESULT 12

64774
ybaW protein Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 24-Nov-1999

C:Accession: E64774

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shaq, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; M010:9742617

A:Accession: E64774

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-132 (BLAT)

A:Cross-references: GH:AF00150; CH:U00096; NID:gl786639; PID:gl786647;

A:Experimental source: Strain K-12, substrain M61655

C:Genetics:

A:Gene: ybaW

C:Superfamily: 15.5K protein (toLAB operon 5' region)

Query Match

Best Local Similarity 100.0%; Score 7; DH 2; Length 132;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 GILSVI 194
IIIIII
DB 84 GILSVI 90

RESULT 13

E85541

hypothetical protein ybaW [imported] - Escherichia coli (strain 0157:H7, substrain EDL93

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: E85541

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, E.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Damalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 524-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; M010:21074935; PMID:11206551

A:Accession: E85541

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-132 (SP)

A:Cross-references: GH:AR005174; NID:gl2513305; PID:AA054793.1; GSPTB:GN00145; UWGP:Z05

A:Experimental source: strain 0157:H7, substrain EDL933

C:Genetics:

A:Gene: ybaW

C:Superfamily: 15.5K protein (toLAB operon 5' region)

Query Match

Best Local Similarity 100.0%; Score 7; DH 2; Length 132;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 GILSVI 194
IIIIII
DB 84 GILSVI 90

RESULT 14

E72603

hypothetical protein APE1292 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C:Accession: E72603

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Baikawa, Y.; Jin-uno, K.; To

awa, H.; Takamiya, M.; Masuda, S.; Furubashi, T.; Tanaka, T.; Kodoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Gram-negative, Aero

A:Reference number: A72450; M01:99310339

A:Accession: E72603

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-132 (KAW)

A:Cross-references: PHJ:AP000031; NID:gl510182.1; PID:HA020283.1; PID:dl044069; PIR

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1292

Query Match

Best Local Similarity 100.0%; Score 7; DH 2; Length 132;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 LAALOKV 174
IIIIII
DB 110 LAALOKV 116

RESULT 15

T47845

hypothetical protein I209.190 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47845

R:Nyakatura, G.; Fartmann, B.; Bauner, D.; Sterr, W.; Holland, K.; Weichselgartner, M

submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24475

A:Accession: T47845

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-138 (NYA)

A:Cross-references: EMBL:AL138658

A:Experimental source: cultivar Columbia; HAT clone 1209

C:Genetics:

A:Map position: 3

A:Introns: 31/1; 50/3; 67/3; 89/3; 105/3; 115/3

A:Note: I209.190

Query Match

Best Local Similarity 100.0%; Score 7; DH 2; Length 138;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 DLLAIVQ 60
IIIIII
DB 132 DLLAIVQ 138

Search completed: July 2, 2002, 16:15:38

Job time: 755 sec

Db 5 KKKRGDPE 12

RESULT 7

C84747
 Probable protein kinase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse cat cross)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02 Feb-2001
 C:Accession: C84747
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shew, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Niernman, W.C.; White, G.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: C84747
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-664 <STO>
 A:Cross-references: GB:AF002093; NID:g2459440; PIDN:AA8806/5.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g33580
 A:Map position: 2

Query Match 4.0%, Score 8; DB 2; Length 664,
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKRGDPE 69
 |||||
 DB 567 KKKRGDPE 574

RESULT 8

AD2443
 hypothetical protein al15100 [imported] - Anabaena sp. (strain PCC 7120)
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14 Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AD2443
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchida,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2443
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1906 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA076799.1; PID:q17134238; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: al15100

Query Match 4.0%, Score 8; DB 2; Length 1906;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 NKLOONIS 39
 |||||
 DB 1595 NKLOONIS 1602

RESULT 9

E97706
 hypothetical protein KC0053 [imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
 C:Accession: E97706
 R:Odada, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Harbo, V.; Samson, D.; R
 Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; MUID:21442074; PMID:11557893
 A:Accession: E97706
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-72 <KUR>
 A:Cross-references: GB:AF006914; PIDN:AA102591.1; PID:g15619080; GSPDB:GN00173
 C:Genetics:
 A:Gene: RC0053

Query Match 3.5%, Score 7; DB 2; Length 72;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 RQTALYT 80
 |||||
 DB 29 RQTALYT 35

RESULT 10

E72707
 hypothetical protein APE1076 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20 Aug-1999
 C:Accession: E72707
 R:Kawarayashi, Y., Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper thermophilic Gram-negative bacterium, Aery
 A:Reference number: A72450; MUID:99310339
 A:Accession: E72707
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-105 <RAW>
 A:Cross-references: DB:AF000009; NID:g5164188; PIDN:AAA60061.1; PID:d1043847; PID:g
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1076

Query Match 3.5%, Score 7; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 PLPSVPR 24
 |||||
 DB 93 PLPSVRR 99

RESULT 11

A90691
 hypothetical protein ECs0497 [imported] - Escherichia coli (strain 0157:H7, substrain
 C:Species: Escherichia coli
 C:Date: 18 Jul-2001 #sequence_revision 18 Jul 2001 #text_change 03 Aug 2001
 C:Accession: A90691
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g
 A:Reference number: A90629; MUID:21156231; PMID:11258796
 A:Accession: A90691
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-132 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BAH33920.1; PID:g1335954; GSPDB:GN00154
 A:Experimental source: strain 0157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: ECs0497
 C:Superfamily: 15.5K protein (G1AR superfamily 5' region)

Query Match 3.5%, Score 7; DB 2; Length 132;

Query Match 4.0% Score 8; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKEEEE 49
 DB 77 KKKKEEEE 84

RESULT 3

S75546
 Hypothetical protein, *Beirococcus radiodurans* (strain P1)
 C:Species: *Beirococcus radiodurans*
 C:Date: 03-Dec-1999 #sequence_revision 03 Dec 1999 #text_change 31 Mar 2000
 C:Accession: J75546
 A:Authors: Eison, D.; Eison, J. A.; Eide, J. E.; Hickey, F. K.; Peterson, J. D.; Dodson, R. L.;
 M.; Shen, M.; Vamathevan, J.; Lam, P.; McDonald, L.; Otterback, T.; Zalewski, C.; Ma
 S.; Smith, B.; Venter, J. C.; Fraser, C. M.
 A:Title: Genome sequence of the radioresistant bacterium *Beirococcus radiodurans* P1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: J75546
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1123 <WHI>
 A:Cross-references: GR:AF001884; GR:AF000513; NID:g6457878; PIDN:AAF09801.1; PID:g645788
 A:Experimental source: strain K1
 A:Genetics:
 A:Gene: DR0207
 A:Map position: 1

Query Match 4.0% Score 8; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 LLSALAAL 171
 DB 5 LLSALAAL 12

RESULT 4

S73484
 Probable lipoprotein K02-07541, *Mycoplasma pneumoniae* (strain AICC 29342)
 C:Species: *Mycoplasma pneumoniae*
 A:Variety: AICC 29342
 C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07 Dec 1999
 C:Accession: S73484
 R:Hummelreich, R.; Hilbert, H.; Plagens, B.; Pirkl, E.; Li, B. C.; Herrmann, R.
 A:Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*
 A:Reference number: S73427; MUID:9710585
 A:Accession: S73484
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1541 <WHI>
 A:Cross-references: EMBL:AF000007; GR:U00089; NID:q1674704; PIDN:AAH95705.1; PID:q167370
 A:Note: The nucleotide sequence was submitted to the EMBL data library, November 1996
 A:Genetics:
 A:Genetic code: SGC3

Query Match 4.0% Score 8; DB 2; Length 541;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 AALQKVE 176
 DB 50 AALQKVE 57

RESULT 5

A70038
 L-lactate permease homolog yvfh - *Bacillus subtilis*
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence_revision 05 Dec 1997 #text_change 20-Jun-2000
 C:Accession: A70038
 A:Authors: Ogasawara, N.; Mester, T.; Albertini, A. M.; Allou, G.; Azabou, V.; Ber
 F.; Kinet, F.; Ogasawara, N.; Mester, T.; Albertini, A. M.; Allou, G.; Azabou, V.; Ber
 C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.;
 A.; Ehrlich, S. D.; Emerson, P. T.; Entian, K. D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 tech, J.; Harwood, C. R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Rullo, M.
 Koether, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S. M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S. H.; Parro, V.; Pohl, T. M.; Portete
 Rieder, M.; Rivolta, C.; Rocha, E.; Roche, R.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
 A:Authors: Schiebel, S.; Schroeder, K.; Scroffone, F.; Sckiguchi, J.; Sekowska, A.; Se
 akouchi, M.; Yamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tostato, V.; Uehiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H. F.; Zumbstein, E.; Yoshikawa, H.; Zumbstein, E.; Zumbstein, E.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: A70038
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-563 <KUN>
 A:Cross-references: GR:260121; GR:AL009125; NID:g260827; PIDN:CAF5424.1; PID:g26089
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yvfh
 C:Superfamily: L lactate permease

Query Match 4.0% Score 8; DB 2; Length 563;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 LLSALAAL 171
 DB 15 LLSALAAL 22

RESULT 6

GB6449
 F5D14.23 protein, *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse ear cross)
 C:Date: 02-Mar-2001 #sequence_revision 02 Mar 2001 #text_change 03 May 2001
 C:Accession: GB6449
 R:Theologis, A.; Ecker, J. R.; Palm, C. J.; Federspiel, N. A.; Kaul, S.; White, T.; Alon
 chin, C. W.; Chung, M. K.; Conn, L.; Conway, A. B.; Green, A. R.; Greasy, T. H.; Lewan,
 ansen, N. F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J. L.; Jenkins, J.; Johnson-Hopson, G.; Khan, S.; Klaykin, E.; Kim,
 C. A.; Li, J. H.; Li, Y.; Lin, X.; Liu, S. X.; Liu, Z. A.; Lopez, J. S.; Maiz, P.; Marz
 Blazzo, M.; Rooney, T.; Rowley, P.; Sakano, H.
 A:Authors: Salberg, S. I.; Schwartz, J. P.; Shinn, P.; Southwick, A. M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C. M.; Venter, J. C.; Davis, R. W.
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; MUID:21016719
 A:Accession: GB6449
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-606 <STO>
 A:Cross-references: GR:AF005172; NID:g9920621; PIDN:AAF0144.1; GSIDB:GN 0141
 C:Genetics:
 A:Map position: 1

Query Match 4.0% Score 8; DB 2; Length 606;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKEEEE 69
 DB 11111111

GenCore version 4.5
Copyright (c) 1993-2000 Genophen Ltd

OM protein protein search, using sw model

Run on: July 2, 2002, 16:03:03 ; Search time 57.09 Seconds
(without alignments)

338,307 Million cell updates/sec

Title: US 09-603-665 5_COPY_1594_1794

Perfect score: 201
Sequence: 1 LPTETETFPVARGVGNPLP... (SPVLEGLTSCV:HLKITS 201

Scoring table: OLIGO
Gapop 60 0 , Gapext 60.0

Searched: 28338 seqs, 96089334 residues

Word size : 6

Total number of hits satisfying chosen parameters: 1351

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :
PIR71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	4.0	108	2 B75029	hypothetical prote
2	8	4.0	123	1 G71214	hypothetical prote
3	8	4.0	123	2 B75546	hypothetical prote
4	8	4.0	541	2 S73383	probable lipoprote
5	8	4.0	563	2 A70038	L-lactate permease
6	8	4.0	606	2 G86449	L5014_23 protein -
7	8	4.0	664	2 G84747	probable protein k
8	8	4.0	1906	2 AD2443	hypothetical prote
9	7	3.5	72	2 E97706	hypothetical prote
10	7	3.5	105	2 E72707	hypothetical prote
11	7	3.5	132	2 A90691	hypothetical prote
12	7	3.5	132	2 G4774	zbow protein Esc
13	7	3.5	132	2 B85541	hypothetical prote
14	7	3.5	132	2 E72603	hypothetical prote
15	7	3.5	148	2 I47445	hypothetical prote
16	7	3.5	143	2 H72610	hypothetical prote
17	7	3.5	146	2 H81840	hypothetical prote
18	7	3.5	146	2 C81100	hypothetical prote
19	7	3.5	178	2 JN0475	interleukin-10 pre
20	7	3.5	180	2 T36927	hypothetical prote
21	7	3.5	214	2 T10737	extensin-like cell
22	7	3.5	214	2 T09854	proline-rich cell
23	7	3.5	216	1 FV0038	colanic acid biosy
24	7	3.5	216	2 B85851	hypothetical prote
25	7	3.5	216	2 A69790	regulator of capsu
26	7	3.5	216	2 B91017	colanic acid biosy
27	7	3.5	217	2 E69271	hypothetical prote
28	7	3.5	217	2 A63149	probable two compo
29	7	3.5	218	2 S50566	hypothetical prote

GMP synthase xpro56
hypothetical prote
hypothetical prote
40S ribosomal prot
probable molybdopt
hypothetical prote
conserved hypotet
phorol protein - yC
huntingtin-associat
huntingtin-associat
c4-dicarboxylate-b
hypothetical prote
hypothetical prote
hypothetical prote
probable membrane
alcohol dehydrogen

ALIGNMENTS

RESULT 1

B75029

hypothetical protein PAB1179 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #Sequence_revision 20-Aug-1999 #Text_change 20-Aug-2000

C:Accession: B75029

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence. Insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: B75029

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1 108 <RAW>

A:Cross-references: CR:AL048288; CR:AL048288; NID:G5458940; PDB:1A9672 1; PDB:G545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1179

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein M0223

Query Match 4.0%; Score 8; DB 2; Length 108;

Best Local Similarity 100.0%; Prod. No. 2, 5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKKEEE 69

DB 62 KKKKKEEE 69

RESULT 2

G71214

hypothetical protein PH983 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 20-Sep-1999 #Sequence_revision 10-Sep-1999 #Text_change 21-Jul-2000

C:Accession: G71214

P:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Onfuku, Y.; Fushishiro, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Ogi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper thermophil

A:Reference number: A71000; M01000; M01000; M01000

A:Accession: G71214

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1 123 <RAW>

A:Cross-references: CR:AP000007; NID:G326124; PDB:1A9672 1; PDB:G326124

A:Experimental source: Strain OT3

A:Note: This accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH983

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein M0223

? NAME: Baumeister, Kirk
? REGISTRATION NUMBER: 33,833
? REFERENCE/DOCKET NUMBER: P31169
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 610-270-5096
? TELEFAX: 610-270-5090
? TELEX:
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 46 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? FEATURE:
? NAME/KEY: Other
? LOCATION: 1...1
? OTHER INFORMATION: Residue 1 methoxy pyrylated
? NAME/KEY: Other
? LOCATION: 22...22
? OTHER INFORMATION: Disulfide linkage to SEQ ID NO: 7
US-09-391-799-8

Query Match 3.0%; Score 6; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0, Mismatches 0, Indels 0, Gaps 0,

OY 42 KLVIVR 47
Db 12 KLVIVR 17

Search completed: July 2, 2002, 16:03:57
Job time: 54 sec

RESULT 13
US-09-557-409B-66
Sequence 66: Application US/09-557-409B
Patent No. 691672
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yash A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Boudreau, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF I
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and HERRY LLP
STREET: 6400 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.3C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09-557-409A
FILING DATE: 14 Nov 1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Mark, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 34 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
US-09-557-409B-66

Query Match 3.0%, Score 6; DB 2; Length 34;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 141 TLEALA 136
Db 12 TLEALA 17

RESULT 14
US-09-391-799-5
Sequence 5: Application US/09-391-799
Patent No. 6187751
GENERAL INFORMATION:
APPLICANT: Smith, Richard
APPLICANT: Reelley, Lee
TITLE OF INVENTION: Biologically Active Peptide
TITLE OF INVENTION: Fragments of Ob Protein
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09-391-799
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Baumelster, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P31169
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Other
LOCATION: 1...1
OTHER INFORMATION: Residue 1 acylated
NAME/KEY: Other
LOCATION: 22...22
OTHER INFORMATION: Disulfide linkage to SEQ ID No: 6
US-09-391-799-5

Query Match 3.0%, Score 6; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 11002;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
QY 42 KTIVTR 47
Db 12 KTIVTR 17

RESULT 15
US-09-391-799-8
Sequence 8: Application US/09-391-799
Patent No. 6187751
GENERAL INFORMATION:
APPLICANT: Smith, Richard
APPLICANT: Reelley, Lee
TITLE OF INVENTION: Biologically Active Peptide
TITLE OF INVENTION: Fragments of Ob Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09-391-799
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

RESULT 10
US-09-391-799-1
; Sequence 1, Application US/09391799
; Patent No. 6187751
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard
; TITLE OF INVENTION: Biologically Active Peptide
; TITLE OF INVENTION: Fragments of OB Protein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09391.799
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 34,833
; REFERENCE/DOCKET NUMBER: P31169
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-391-799-1

Query Match 3.0%, Score 6; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 KTIIVR 47
Db 12 KTIIVR 17

RESULT 11
US-09-961-749-7
; Sequence 7, Application US/08961749
; Patent No. 5908830
; GENERAL INFORMATION:
; APPLICANT: SMITH, ROY G.
; APPLICANT: CASCIELLO, MARGARET A.
; APPLICANT: MACINTYRE, RUAN
; APPLICANT: MACNEILL, DOUGLAS J.
; APPLICANT: MENKE, JOHN G.
; TITLE OF INVENTION: COMBINATION THERAPY FOR THE TREATMENT OF
; TITLE OF INVENTION: DIABETES AND OBESITY
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MARY A APOLLINA - MERCK & CO., INC.
; STREET: 122 EAST LINCOLN AVENUE - P.O. Box 2000
; CITY: RAHWAY

STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1 0, Version #1 30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,749
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: APOLLINA, MARY A.
REGISTRATION NUMBER: 34,087
REFERENCE/DOCKET NUMBER: 19822Y
TELECOMMUNICATION INFORMATION:
TELEPHONE: (732) 594-3462
TELEFAX: (732) 594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-749-7

Query Match 3.0%, Score 6; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 KTIIVR 47
Db 15 KTIIVR 20

RESULT 12
US-09-461-697-281
; Sequence 281, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 281
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-281

Query Match 3.0%, Score 6; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 LQKVVE 176
Db 11 LQKVVE 16

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 883 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US 08 596-466-2

Query Match 3.5%, Score 7; DH 2; Length 883;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 ALQKVE 176
 DB 737 ALQKVE 743

RESULT 8
 US-08-967-104-2
 : Sequence 2, Application US/08967104
 : Patent No. 5919694
 : GENERAL INFORMATION:
 : APPLICANT: SUGIMOTO, Masakazu
 : APPLICANT: SUZUKI, Tomoko
 : APPLICANT: MATSUI, Hiroshi
 : APPLICANT: IZUI, Katsura
 : TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE,
 : TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
 : NUMBER OF SEQUENCES: 12
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
 : ADDRESSEE: P.C.
 : STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
 : CITY: ARLINGTON
 : STATE: VIRGINIA
 : COUNTRY: USA
 : ZIP: 22202
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/967,104
 : FILING DATE:
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/596,466
 : FILING DATE: 29 APR-1996
 : APPLICATION NUMBER: JP 5,209775
 : FILING DATE: 24 AUG-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 5-209776
 : FILING DATE: 24-AUG-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 6-153876
 : FILING DATE: 05-JUL-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: OBLON, NORMAN F.
 : REGISTRATION NUMBER: 24,618
 : REFERENCE/DOCKET NUMBER: 10-784-0 PCT
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 413-3000
 : TELEFAX: (703) 413-2220
 : TELEX: 248855 OPAT DR
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 883 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-967-104-2

Query Match 3.5%, Score 7; DH 2; Length 883;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 ALQKVE 176
 DB 737 ALQKVE 743

RESULT 9
 US-08-159-339A-312
 : Sequence 312, Application US/08159339A
 : Patent No. 6037135
 : GENERAL INFORMATION:
 : APPLICANT: Kubo, Ralph T.
 : APPLICANT: Grey, Howard M.
 : APPLICANT: Sette, Alessandro
 : APPLICANT: Cellis, Esteban
 : TITLE OF INVENTION: HLA Binding peptides and their
 : TITLE OF INVENTION: Uses
 : NUMBER OF SEQUENCES: 1254
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend and Crew LLP
 : STREET: Two Embarcadero Center, Eighth Floor
 : CITY: San Francisco
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94111-3834
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: PastSeq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/159,339A
 : FILING DATE: 29-NOV-1993
 : CLASSIFICATION: 424
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/926,666
 : FILING DATE: 07-AUG-1992
 : APPLICATION NUMBER: US 08/047,746
 : FILING DATE: 05-MAR-1993
 : APPLICATION NUMBER: US 08/103,396
 : FILING DATE: 06-AUG-1993
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Weber, Ellen Lauver
 : REGISTRATION NUMBER: 32,762
 : REFERENCE/DOCKET NUMBER: 018623-0050300S
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 576-0200
 : TELEFAX: (415) 576-0300
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 312:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 10 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: peptide
 : US-08-159-339A-312

Query Match 3.0%, Score 6; DH 3; Length 10;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 ALAIPQ 139
 DB 3 ALAIPQ 8

Query Match 3.5%; Score 7; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 164 LLSALAA 170
Db 56 LLSALAA 62
|||||||

RESULT 5
US-08-770-035-1
; Sequence 1, Application US/08770035
; Patent No. 6008342
; GENERAL INFORMATION:
; APPLICANT: Ringer, Mary-Helen
; APPLICANT: Pasamontes, Luis
; TITLE OF INVENTION: Coccidioides Vaccines
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07110

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/729,099

; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Eimeria tenella
; DEVELOPMENTAL STAGE: merozoite

US-08-770-035-1

Query Match 3.5%; Score 7; DB 3; Length 315;
Best Local Similarity 100.0%; Pred No 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 164 LLSALAA 170
Db 56 LLSALAA 62
|||||||

RESULT 6
US-08-556-419-25
; Sequence 25, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang

; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Ianahan, Anthony
; APPLICANT: Worley, Paul
; APPLICANT: Snyder, Solomon
; TITLE OF INVENTION: Huntingtin-associated protein
; FILE REFERENCE: 01107 52271
; CURRENT APPLICATION NUMBER: US/08/556,419C
; CURRENT FILING DATE: 1995-11-09
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-556-419-25

Query Match 3.5%; Score 7; DB 3; Length 331;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 VTSTLEA 134
Db 307 VTSTLEA 313
|||||||

RESULT 7
US-08-596-366-2
; Sequence 2, Application US/08596366
; Patent No. 5876983
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE,
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596,366
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209775
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209776
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-153876
; FILING DATE: 05-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: ORLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-784-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Charles and Brady
 STREET: P.O. Box 2113
 CITY: Madison
 STATE: WI
 COUNTRY: USA
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/217,327
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/812,233
 FILING DATE: 19-DEC-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Soay, Nicholas J.
 REGISTRATION NUMBER: 27,386
 REFERENCE/DOCKET NUMBER: 1122990831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9186
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 214 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-217-327-4

Query Match 3.5%, Score 7; DB 1; Length 214;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 SALLCIA 126
 DB 11 SALLCIA 17

RESULT 3
 US-07-729,099-1
 Sequence 1, Application US/07729099
 Patent No. 5403581
 GENERAL INFORMATION:
 APPLICANT: Bingner, Mary-Helen
 APPLICANT: Pasamontes, Luis
 TITLE OF INVENTION: Coccidiosis Vaccines
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hoffmann-La Roche Inc.
 STREET: 340 Kingsland Street
 CITY: Nutley
 STATE: New Jersey
 COUNTRY: U.S.
 ZIP: 07110
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/729,099
 FILING DATE: 19910712
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Roseman, Catherine R
 REGISTRATION NUMBER: 34,240
 REFERENCE/DOCKET NUMBER: 8514

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (201) 235-6208
 TELEFAX: (201) 235-3500
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 315 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Elmeria tenella
 DEVELOPMENTAL STAGE: merozoite
 US 07 729-099-1

Query Match 3.5%, Score 7; DB 1; Length 315;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 LLSALIAA 170
 DB 56 LLSALIAA 62

RESULT 4
 US-08-257-392-1
 Sequence 1, Application US/08:57392
 Patent No. 5688513
 GENERAL INFORMATION:
 APPLICANT: Bingner, Mary-Helen
 APPLICANT: Pasamontes, Luis
 TITLE OF INVENTION: Coccidiosis Vaccines
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hoffmann-La Roche Inc.
 STREET: 340 Kingsland Street
 CITY: Nutley
 STATE: New Jersey
 COUNTRY: U.S.
 ZIP: 07110
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/257,392
 FILING DATE: 09-JUN-1994
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/729,099
 FILING DATE: 12-JUL-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Roseman, Catherine R
 REGISTRATION NUMBER: 34,240
 REFERENCE/DOCKET NUMBER: 8514
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (201) 235-6208
 TELEFAX: (201) 235-3500
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 315 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Elmeria tenella
 DEVELOPMENTAL STAGE: merozoite
 US-08-257-392-1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model
Run on: July 2, 2002, 16:33:03, Search time: 41:05 seconds
(without alignments)
100 466 Million cell updates/sec

Title: US-09-603-665-5_COPY_1594_1794
Perfect score: 201
Sequence: 1 LAPTETP(PVIRGI)GNP.P.....ISPVLEGIISQVHLEKITS 201

Scoring table: OLIGO
Gap: 60 0 0, Gap-ext: 60 0

Searched: 231628 seqs, 24425594 residues

Word size: 6

Total number of hits satisfying chosen parameters: 387

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database:
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5H_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	4.0	829	4	US 09-413-814-105
2	7	3.5	214	1	US 08-217-327-4
3	7	3.5	315	1	US-07-729-099-1
4	7	3.5	315	1	US-08-257-192-1
5	7	3.5	315	3	US-08-770-035-1
6	7	3.5	331	3	US-08-556-419-25
7	7	3.5	883	2	US-08-596-366-2
8	7	3.5	883	2	US-08-967-104-2
9	6	3.0	10	3	US-08-159-339A-312
10	6	3.0	18	4	US-09-391-799-1
11	6	3.0	20	2	US 08 961 749-7
12	6	3.0	32	4	US 09-461-697-281
13	6	3.0	34	2	US-08-557-309B-66
14	6	3.0	46	4	US-09-391-799-5
15	6	3.0	46	4	US-09-391-799-8
16	6	3.0	48	2	US-08-347-563A-23
17	6	3.0	48	3	US-08-485-942A-23
18	6	3.0	48	4	US-08-488-214A-23
19	6	3.0	48	4	US-08-488-208A-23
20	6	3.0	48	4	US-08-483-211A-23
21	6	3.0	48	4	US 08-488-227A-23
22	6	3.0	63	2	US-08-466-583A-8
23	6	3.0	63	5	PCF-US95-07820-8
24	6	3.0	74	4	US-08-858-207A-409
25	6	3.0	76	4	US-08-805-940-5
26	6	3.0	79	2	US 08 839 709-3
27	6	3.0	79	2	US-08-839-709-4

ALIGNMENTS

RESULT 1
US-09-413-814-105
; Sequence 105, Application US/09413814
; Patent No. 5225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Heyer, Stefan
; APPLICANT: Hoescker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hufle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; TITLE OF INVENTION: heteropolypeptide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 105
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-105

Query Match 4.0%; Score 8; DB 4; length 829;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 164 LLSAIAAL 171
DB 309 LLSAIAAL 316
RESULT 2
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
; APPLICANT: John, Malliyakal E
; APPLICANT: Barton, Kenneth A
; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
; NUMBER OF SEQUENCES: 16


```

; NUMBER OF SEQ ID NOS: 2660
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2541
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (79)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (100)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-757 028 2541

Query Match
Best Local Similarity 13.4%; Score 27; DB 21; Length 105;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 VETLPHFISYLEGILSQVHLEKITS 201
Db 19 VETLPHFISYLEGILSQVHLEKITS 45

RESULT 13
US-09-248-796 20694
; Sequence 20694, Application US/09/248736
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107106-132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20694
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-20694

Query Match
Best Local Similarity 5.0%; Score 10; DB 16; Length 241;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SVRRKALDL 30
Db 89 SVRRKALDL 98

RESULT 14
US 60 096 409 20694
; Sequence 20694, Application US/000454094
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 010 016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20694
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Candida albicans
```

US-60-096-409-20694

```

Query Match
Best Local Similarity 5.0%; Score 10; DB 26; Length 241;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SVRRKALDL 30
Db 89 SVRRKALDL 98

RESULT 15
PCT-US01-08631-32819
; Sequence 32819, Application PCT/US0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 32819
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-32819

Query Match
Best Local Similarity 4.0%; Score 8; DB 1; Length 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKEGEE 69
Db 29 KKKEGEE 36

Search completed: July 2, 2002, 16:14:34
Job time: 691 sec
```

```

; NAME/KEY: DOMAIN
; LOCATION: (476)..(427)
; OTHER INFORMATION: case12s alpha/beta proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number R100406B, p-value=7.309e-09, raw score of 8.28
PCT-US01-086631-47591

```

Query Match	30.88;	Score	62;	DB	1;	Length	1569;
Best Local Similarity	100.0%;	Fred. No.	2.6e-51;				
Matches	62;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;

Q7 1 LLLPTETFLPVIRGLVGNPLPSVRRKALDIANNKIQONISWKKTVTRFLKLVDPDIAIVQ 60
|||||
Db 1508 LLLPTETFLPVIRGLVGNPLPSVRRKALDIANNKIQONISWKKTVTRFLKLVDPDIAIVQ 1567

Qy 61 RK 62

—

RESULT 9
USCG 757 028 1640
Sequence 1640 Application US/09757028
GENERAL INFORMATION:
APPLICANT: ROSEN, ET AL.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PM001
CURRENT APPLICATION NUMBER: US/09757,028

Query Match	26.4%	Score 53:	DB 21:	Length 170:
Best local Similarity	100.0%	Pred. No. 2.4e-43:		
Matches 53:	Conservative	0:	Mismatches	0:
	Indels	0:	Gaps	0:

QY 48 FLKLPDLLAIIV₉KKKKKE₁₀EEEE₁₁AIN₁₂Q₁₃IAI₁₄YTLKLL₁₅KNF₁₆;AENPDPFVPVL 100
 |||||
 DB 42 FLKLPDLLAIIV₉KKKKKE₁₀EEEE₁₁AIN₁₂Q₁₃IAI₁₄YTLKLL₁₅KNF₁₆GAENPDPFVPVL 94
 |||||

RESULT 10
US-09-758 449 1058
; Sequence 1058, Application US/09758449
; GENERAL INFORMATION:
; APPLICANT: ROSEN et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

: FILE REFERENCE: PM026.
: CURRENT APPLICATION NUMBER: US/09/758.449
: PRIORITY FILING DATE: 2001-01-11
: PRIOR APPLICATION NUMBER: 60/179,065
: PRIOR FILING DATE: 2000-01-31
: PRIOR APPLICATION NUMBER: 60/180,628
: PRIOR FILING DATE: 2000-02-04
: NUMBER OF SEQ ID NOS: 1478
: SOFTWARE Patent.In Ver. 2.0
: SEQ ID NO 1058
: LENGTH: 164
: TYPE: prt
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (80)
: OTHER INFORMATION: xaa equals any of the
: US-09-758-449-1058

```

Query Match	23.9%	Score 48	DB 21	Length 164
Rest Local Similarity	100.0%			
Matches	48	Constractive	0	Indels
		Mismatches	0	Gaps

[illegible]

```

RESULT 11
US-04-757-028-2590
Sequence 2590, Application US/09757028
GENERAL INFORMATION:
APPLICANT: P. Sen et al
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PM001
CURRENT APPLICATION NUMBER: US/09/757,028
CURRENT FILING DATE: 2001-01-09
PRIORITY APPLICATION NUMBERS: 60/179,065
PRIORITY FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
NUMBER OF SEQ ID NOS: 2660
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2590
LENGTH: 135
TYPE: PRT
ORGANISM: Homo sapiens
US-09 757-028-2590

```

Query Match.	14 4%	Score	20	EB	21	Length	135
Best Local Similarity	100 0%	Pred No.	9a	20			
Matches	29	Conservative	0	Mismatches	0	Indels	0
Gaps	0						

QY 173 KVVETLPHFISPYLEGLISQVIHLKITS 201
| | | | | | | | | | | | | | | |
DB 44 KVVETLPHFISPYLEGLISQVIHLKITS 72

```

RESULT 12
US-09-757-028-2541
; Sequence 2541, Application US/09757028
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM001
; CURRENT APPLICATION NUMBER: US/09/757,028
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04

```

STRANDEDNESS:
TOPOLGY: linear
US 08 936 487-52

Query Match 47.8%, Score 76, DB 13, Length 515;
Best Local Similarity 100.0%, Pred. No. 1.4e-65;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 185
DB 90 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 149
QY 186 LEGILSQVHLEKITS 201
DB 150 LEGILSQVHLEKITS 165

RESULT 6

US 60-042-985-52
Sequence 52, Application 95/50042985
GENERAL INFORMATION:
APPLICANT: Howcock, Anne M.
APPLICANT: Baer, Richard
TITLE OF INVENTION: Compositions and Methods Comprising
TITLE OF INVENTION: BARD1 and other BRCA1 Binding Proteins
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: Concurrently Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W.
REGISTRATION NUMBER: P-41,071
REFERENCE/DOCKET NUMBER: USID:499P23
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 515 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-60-042-985-52

Query Match 47.8%, Score 76, DB 26, Length 515;
Best Local Similarity 100.0%, Pred. No. 1.4e-65;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 185
DB 90 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 149
QY 186 LEGILSQVHLEKITS 201
DB 150 LEGILSQVHLEKITS 165

Query Match 47.8%, Score 76, DB 26, Length 515;
Best Local Similarity 100.0%, Pred. No. 1.4e-65;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 185
DB 90 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 149
QY 186 LEGILSQVHLEKITS 201
DB 150 LEGILSQVHLEKITS 165

RESULT 7

US-60-042-985-52
Sequence 52, Application US/50042985
GENERAL INFORMATION:
APPLICANT: Howcock, Anne M.
APPLICANT: Baer, Richard
TITLE OF INVENTION: Compositions and Methods Comprising
TITLE OF INVENTION: BARD1 and other BRCA1 Binding Proteins
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: Concurrently Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W.
REGISTRATION NUMBER: P-41,071
REFERENCE/DOCKET NUMBER: USID:499P23
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 515 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-60-042-985-52

Query Match 47.8%, Score 76, DB 26, Length 515;
Best Local Similarity 100.0%, Pred. No. 1.4e-65;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 185
DB 90 AEVSTLEALAIPOIPLSIMPSTLITMKNTSELVSEVYLLSALAALOKVVTLPPIFISPY 149
QY 186 LEGILSQVHLEKITS 201
DB 150 LEGILSQVHLEKITS 165

RESULT 8

PCI-US01-08631-45591
Sequence 45591, Application PC/1050108631
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 232/2-049
CURRENT APPLICATION NUMBER: PCI/US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 45591
LENGTH: 1569
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:


```

: LOCATION: 1967
: OTHER INFORMATION: Xaa-Asp or Asn
: NAME/KEY: VARIANT
: LOCATION: 2017
: OTHER INFORMATION: Xaa-Gly or Glu
US-09-603-665-5

Query Match
Best Local Similarity 99.5%, Score 200, DB 20, Length 2144,
Matches 201, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1 LLPTETFPVIRGLVGNPLSPVRRKALDNLNNKLOONISWKKTIIVTFKLKVPDLIAIVQ 60
DB 1594 LLPTETFPVIRGLVGNPLSPVRRKALDNLNNKLOONISWKKTIIVTFKLKVPDLIAIVQ 1653
QY 61 RKKKEGFEQAINRQIALYTIKILCKNFGAENDDPVVIXIAVKLIAPRKEKNVLGS 120
DB 1654 RKKKEGFEQAINRQIALYTIKILCKNFGAENDDPVVIXIAVKLIAPRKEKNVLGS 1713
QY 121 ALLCIAEVTSTLEALAIPOQLSPMSLITMKNTSELVSSEVYLLSALAALOKVVTETLPH 180
DB 1714 ALLCIAEVTSTLEALAIPOQLSPMSLITMKNTSELVSSEVYLLSALAALOKVVTETLPH 1773
QY 181 FISPYLEGILSOVHLEKITS 201
DB 1774 FISPYLEGILSOVHLEKITS 1794

RESULT 3
US-60-141-323-3
: Sequence 3, Application US/60141323
: GENERAL INFORMATION:
: APPLICANT: Barry, Caroline
: APPLICANT: Houqueleret, Lydie
: APPLICANT: Chumakov, Ilya
: TITLE OF INVENTION: A NOVEL HAP28 GENE AND PROTEIN
: FILE REFERENCE: GENSET 063PRF
: CURRENT APPLICATION NUMBER: US/60/141,323
: EARLIER FILING DATE: 1999-06-25
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: Patent.pm
: SEQ ID NO 3
: LENGTH: 2144
: TYPE: PRT
: ORGANISM: homo sapiens
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 91..111
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 212..232
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 700..720
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 861..881
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 884..904
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 911..931
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 937..957
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 1297..1317
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 1410..1430
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: 1725..1745
: OTHER INFORMATION: potential membrane-spanning segment, TopPred Program
: FEATURE:

```



```

Best Local Similarity: 100.0%; Prod. No. 48;
Matches: 7; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 61 RKKKEGE 67
DB 18 RKKKEGE 24

RESULT 14
H0 AAM67101 standard; Protein: 42 AA.
XX
AA AAM67101;
XX
XX 06 NOV 2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27407.
XX
XX Human bone marrow expressed exon gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma.
XX
XX Homo sapiens
XX
XX W0200157276 A2.
XX
XX 09 AUG 2001.
XX
XX 40 JAN 2001; 2001WO 0500668.
XX
XX 04 FEB 2000; 2000US 0180312.
XX
XX 26 MAY 2000; 2000US 0207456.
XX
XX 30 JUN 2000; 2000US 0608408.
XX
XX 03 AUG 2000; 2000US 0642356.
XX
XX 21 SEP 2000; 2000US 0234687.
XX
XX 27 SEP 2000; 2000US 0234687.
XX
XX 04 OCT 2000; 2000US 0234687.
XX
XX (MOLE) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow.
XX
XX Claim 27; SEQ ID NO 19788; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENs). See AAM10068 AA:28479. The present sequence is a peptide encoded
XX by one such probe. The SENs are derived from human cells. The SENs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at http://wipo.int/pub/published\_pct\_sequences.
XX
XX Sequence 42 AA:

Query Match: 3.5%; Score 7; DB 22; Length 42;
Best Local Similarity: 100.0%; Prod. No. 48;
Matches: 7; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 61 RKKKEGE 67
DB 18 RKKKEGE 24

Search completed: July 2, 2002, 16:06:30
Job time: 178 sec

```

PI Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 PS Claim 27; SEQ ID No. 465/4; 65ppp - sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver, the present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 42 AA;

Query Match 3.5%; Score 7; DB 22; Length 42;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 RKKKKEG 67
 Db 18 rkkkeqe 24

RESULT 12

ABBI9374
 ID ABB19374 standard; Protein; 42 AA.

XX AC ABB19374;
 XX

XX 23-JAN-2002 (first entry)
 XX

DE Protein #1373 encoded by probe for measuring heart cell gene expression.
 XX

XX Human, gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX

OS homo sapiens.
 XX

XX WO200157274-A2
 PN

XX 09-AUG-2001
 ID

XX 30-JAN-2001; 2001WO-US00666
 PF

XX 04-FEB-2000; 2000US-0180312
 PR

XX 26-MAY-2000; 2000US-0207456
 PR

XX 30-JUN-2000; 2000US-0608408
 PR

XX 03-AUG-2000; 2000US-0632366
 PR

XX 21-SEP-2000; 2000US-0234687
 PR

XX 27-SEP-2000; 2000US-0236359
 PR

XX 04-OCT-2000; 2000GB-0024263
 PR

XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA

XX Penn SG, Hanzel DK, Chen W, Pank DP;
 PI

XX WPI; 2001-488899/53.
 DR

XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 PT

XX Claim 15; SEQ ID No 21144; 53ppp; English.
 PS

XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535 ABA411305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 42 AA;

Query Match 3.5%; Score 7; DB 22; Length 42;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 RKKKKEG 67
 Db 18 rkkkeqe 24

RESULT 13

AAM54701
 ID AAM54701 standard; Protein; 42 AA.

XX AC AAM54701;
 XX

XX 05-NOV-2001 (first entry)
 XX

DE Human brain expressed single exon probe encoded protein SEQ ID No: 26806.
 XX

XX Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX

OS Homo sapiens.
 XX

XX WO200157275-A2.
 PN

XX 09-AUG-2001
 ID

XX 30-JAN-2001; 2001WO-US00667.
 PF

XX 04-FEB-2000; 2000US-0180312
 PR

XX 26-MAY-2000; 2000US-0207456
 PR

XX 30-JUN-2000; 2000US-0608408
 PR

XX 03-AUG-2000; 2000US-0632366
 PR

XX 21-SEP-2000; 2000US-0234687
 PR

XX 27-SEP-2000; 2000US-0236359
 PR

XX 04-OCT-2000; 2000GB-0024263
 PR

XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA

XX Penn SG, Hanzel DK, Chen W, Pank DP;
 PI

XX WPI; 2001-488899/53.
 DR

XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 PT

XX Example 4; SEQ ID No: 26806; 65ppp + Sequence Listing; English.
 PS

CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 XX

SO Sequence 42 AA;

Query Match

3.5%; Score 7; DB 22; Length 42;

PI responsible for genetic disorders or other traits and to assess
PI biodiversity

PS claim 20; SEQ ID NO 31113; 104pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (I) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC the polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 505 AA:

Query Match 4.0% Score 8; DH 22; Length 505;

Best Local Similarity 100.0%; Prod. No. 53;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LDLLNNKL 44
IIIIIII
DB 248 ldllnnkl 295

RESULT 10

ABB28751
ID ABB28751 standard; Peptide: 42 AA.

XX ABB28751;

DT 01-FEB-2002 (first entry)

DE Peptide #1402 encoded by breast cell single exon nucleic acid probe.

XX Human microarray: single exon probe; gene expression: breast;
KW disease: cancer.

XX Homo sapiens.

XX W0200157271 A2.

XX 09 AUG-2001.

XX 30 JAN 2001; 2001WO-0500662.

XX 04 FEB 2000; 2000US-0180312.

XX 26 MAY 2000; 2000US-0207456.

XX 30 JUN 2000; 2000US-0608408.

XX 03 AUG 2000; 2000US-0642466.

XX 21 SEP 2000; 2000US-0244687.

XX 27 SEP 2000; 2000US-0246359.

XX 04 OCT-2000; 2000US-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,
PI useful for measuring gene expression in sample derived from human
PI breast, comprises number of single exon nucleic acid probes

PS Claim 27; SEQ ID NO 11719; 327pp; sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 42 AA:

Query Match 3.5% Score 7; DH 22; Length 42;

Best Local Similarity 100.0%; Prod. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 RKKKEGE 67
IIIIIII
DB 18 rkkkege 24

RESULT 11

ABB33938

ID ABB33938 standard; Peptide: 42 AA.

XX ABB33938;

DT 04-FEB-2002 (first entry)

DE Peptide #1444 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX W0200157277-A2.

XX 09 AUG-2001.

XX 30 JAN 2001; 2001WO-0500665.

XX 04-FEB-2000; 2000US-0180312.

XX 26 MAY 2000; 2000US-0207456.

XX 30 JUN 2000; 2000US-0608408.

XX 03 AUG 2000; 2000US-0632366.

XX 21 SEP 2000; 2000US-0234687.

XX 27 SEP 2000; 2000US-0246359.

XX 04 OCT-2000; 2000US-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

DR N-PSDB; AAC78035.
 XX Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer
 XX
 PS Claim 11; page 1904-1906; 2352pp; English.
 XX
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given
 CC in AAB43498 to AAB44439. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnervary; immunomodulatory;
 CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
 CC antiinflammatory; antithyroid; anti allergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
 CC neurologic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 363 AA;

Query Match 4.0%; Score 8; DB 21; Length 363;
 Best Local Similarity 100.0%; Pred No 40;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 27 LLLLNKKL 34
 |||||
 Db 146 ldllnkkl 153

RESULT 8
 ID ARG27718 standard; Protein; 435 AA.
 XX
 AC ARG27718;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #27709.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 WIPI: 2001-639362/73.
 DR N-PSDB; AAS91905.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 20; SEQ ID No 58077; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligonucleotides, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement (II) and its binding partners are useful in medical
 CC imaging of cells expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: the sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/pat/published_pat_sequences.
 XX
 SQ Sequence 435 AA;

Query Match 4.0%; Score 8; DB 22; Length 435;
 Best Local Similarity 100.0%; Pred No 47;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 KKKEGEPE 69
 |||||
 Db 396 kkegeccc 403

RESULT 9
 ARG00754
 ID ABG00754 standard; Protein; 505 AA.
 XX
 AC ABG00754;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #745.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 WIPI: 2001-639362/73.
 DR N-PSDB; AAS64941.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations

PI Querellon J, Weissenbach J, Saurin W, Heilla R;
 XX WPI: 2001-126246/14.
 XX
 XX New nucleotide sequences isolated from *Pyrococcus abyssus* encode
 PT proteins useful in industry.
 XX
 XX claim 7: Page 1130: 1637pp: French.
 XX
 XX The present invention relates to the genomic sequence of *Pyrococcus*
 CC abyssus (see AAP8643) and AAH41224-7) and *P. abyssus* proteins. *P. abyssus* is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such *P. abyssus* protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: this patent is in the same patent family as WO200065062, which
 CC contains additional sequences as shown in AAR69132-AA69144,
 CC AAH75903-AAH75920 and AAG66446.
 XX
 XX Sequence 108 AA:

Query Match 4.0%; Score 8; DB 22; Length 138,
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKGEFE 69
 DB | | | | | | | |
 62 kkkkgeee 69

RESULT 6
 AAU14147
 ID AAU14147 standard; Protein: 307 AA.
 XX
 XX AAU14147;

DI 24 OCT-2001 (first entry)
 XX
 XX Human novel protein #8.

XX Human; novel protein; Antianemic; osteopathic; anti-inflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; motropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiatheric; dermatological; haemostatic; antisthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis,
 KW tissue regeneration; immune disorder.

XX Homo sapiens.

XX W200155417 A2.

XX 02 AUG-2001.

XX 25 JAN 2001; 2001WO-052523.

XX 25 JAN 2000; 2000US-0491404.

XX (HYSE) HYSEQ INF.

XX Land YL, Liu C, Dembae RT;

XX WPI: 2001-451949/48.

XX N FSR: AAS22442.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -

XX Example 4; Page 528; 894pp; English.

XX The invention relates to polynucleotides encoding novel human

CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour. In assays to determine biological activity, to
 CC raise antibodies/elicit an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.

XX
 SQ Sequence 307 AA:

Query Match 4.0%; Score 8; DB 22; Length 307;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 IDLLNNKL 34
 DB | | | | | | | |
 90 idllnnkl 97

RESULT 7

AAH43826
 ID AAH43826 standard; Protein: 363 AA.

XX
 XX AAH43826;

XX 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO.1271.

XX Human; cancer associated gene, cancer antigen, detection, cancer;
 KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
 KW antidiabetic; antithyroid; antirheumatic; antiarthritic; antiviral;
 KW antinflamatory; antihypertensive; antiallergic; antitubercular; cardiac;
 KW dermatological; neuroprotective; thrombolytic; coagulant; motropic;
 KW vasotropic; antipsoriatic; antidiabetic; gene therapy; inflammatory;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.

XX Homo sapiens.

XX W6200055350-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05882.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-587533/55.

DR WPI: 2001-639362/73.
 DR N-PSDB: AAS79419.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20: SEQ ID No 45501; 103pp; English
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence: 1569 AA;

Query Match 30.8%; Score 62; DB 22; Length 1569;
 Best Local Similarity 100.0%; Pred. No. 6e-49;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPTETTFIPVHIGLVGNPLPSVRRKALDILNNKIQONI SWKKTIVTRFKLVPHOLIAIVO 60
 DB 1508 LLPTETTFIPVIRGLVGNPLPSVRRKALDILNNKIQONISWKKTLIVTRFKLVPHOLIAIVQ 1567
 QY 61 RK 62
 DB 1568 rk 1569

RESULT 4
 AAG02460
 ID AAG02460 standard; Protein; 68 AA.

XX AAG02460;

XX 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #2451.

KW Human; chromosome mapping; gene mapping; gene therapy; forensics;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-0508631.

XX 31-MAR-2000; 2000US-0543217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
 PT

XX WPI, 2001-639362/73
 DR N-PSDB; AAS66647.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX Claim 20: SEQ ID No 32819; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 68 AA;

Query Match 4.0%; Score 8; DB 22; Length 68;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 KKKKEEEE 69
 DB 29 kkkkeccc 36

RESULT 5

AAB96435

ID AAB96435 standard; Protein; 108 AA.

XX AAB96435;

XX 29-OCT-2001 (first entry)

DE Putative P. abyssii pseudouridine synthase #3.

KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX Pyrococcus abyssi.

XX FR2792651-A1.

XX 27-OCT-2000.

XX 21-APR-1999; 99FR-0005034.

XX 21-APR-1999; 99FR-0005034.

PA (CNES) CNRS CENT NAT RECH SCI
 PA (IPRE-) IPREMER INST FR RECH EXPL MER.

PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX Barry C., Bouquellet L., Chumakov L., Cohen-Akentine A;
 PI WPI: 2001-467032/38.
 XX N PSDB: AAF84907, AAF84910.
 DR
 XX New HAP28 polynucleotides and polypeptides overexpressed in prostate
 PT cancer cells for diagnosing prostate tumors, e.g. by hybridization or
 PI polymerase chain reaction assays.
 XX
 PS Claim 14; Page 297-304; 449pp; English.
 XX
 CC The invention is directed to HAP28 polypeptides, HAP28 polynucleotide
 CC sequences and regulatory region located at the 3' and 5' ends of the
 CC HAP28 coding region, the HAP28 polypeptides can be expressed by standard
 CC recombinant methodology. HAP28 polynucleotides and polypeptides have been
 CC found to be over expressed in prostate tumor cells, therefore levels of
 CC HAP28 expression and/or activity may be assayed (e.g., by polymerase chain
 CC reaction (PCR)) to diagnose patient suffering from or susceptible to
 CC prostate cancer. Antibodies specific for the HAP28 polypeptides are
 CC useful as diagnostic reagents. Allelic markers of the HAP28 gene are
 CC useful in genetic analysis. The present sequence represents a protein
 CC encoded by a first cDNA sequence of the HAP28 gene consisting of the
 CC exons 1 to 45.
 XX
 SQ Sequence: 2144 AA:
 Query Match 99.5%, Score 200; DB 22; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 1,4e-177;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LPTPTFTPTVIGVGVNIPSPVRRALDILNNKIQNLSWKKIIVIRFKIIVNLLAIW 60
 Yb 1594 LPTPTFTPTVIGVGVNIPSPVRRALDILNNKIQNLSWKKIIVIRFKIIVNLLAIW 60
 QY 61 PPKKPK 120
 Yb 1654 PPKKPK 120
 QY 121 ALGTAETVETSTFAIAVQVETPMQSHETPMNSETVSEVYVTSALANCKVYETD 180
 Yb 1714 ALGTAETVETSTFAIAVQVETPMQSHETPMNSETVSEVYVTSALANCKVYETD 180
 QY 181 FLSPLYEGHLSQVHLEKITS 201
 Yb 1774 FLSPLYEGHLSQVHLEKITS 174
 RESULT 2
 AAW54099
 ID AAW54099 standard; Protein: 515 AA.
 AC AAW54099;
 DE 28 SEP 1998 (first entry)
 KW Homo sapiens HAP28 sequence.
 OS Homo sapiens.
 XX
 KW BARD1: ring protein; BRCA1; breast cancer; risk; diagnosis.
 OS Homo sapiens.
 XX
 UN W09812427 A2.
 ID 26 MAR 1998.
 XX
 EF 19-SEP-1997; 97WO-0516842
 PR 04 APR 1997; 97US-0642985.
 PR 20 SEP 1996; 96US-0025296.
 PR 03 APR 1997; 97US-0042611.
 XX

(TEXA) UNIV TEXAS SYSTEM.
 PA Baer R., Howcock AM;
 PI WPI: 1998-230317/20.
 DR N PSDB: AAV24135.
 XX
 CC DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445, which
 CC as breast cancer antigen, BRCA1, binding proteins are useful to
 CC identify patient having or at risk of developing cancer
 CC
 PS Disclosure: Page 287-288; 348pp; English.
 XX
 CC The sequence is that of a protein which can be used in the
 CC preparation of the recombinant breast cancer antigen, BRCA1, binding
 CC proteins BARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the
 CC detection of a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid
 CC sequence, specifically a wild type BARD1 composition for the detection
 CC or purification of BRCA1, useful to identify a patient having, or at
 CC risk of developing cancer. BARD1 can be used in the preparation of an
 CC anti BARD1 antibody, and in the detection and purification of a BRCA1
 CC protein. BARD1, B123, BE2, BE14, BE31 or BE445 can be used in the
 CC identification of a binding protein, agonist or antagonist that alters
 CC the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or the
 CC biological activity of the BRCA1-BARD1, B123, BE2, BE14, BE31 or BE445
 CC complex. The antibodies can be used to detect BARD1, B123, BE2, BE14,
 CC BE31 or BE445, a specific anti-BARD1 antibody can be used to identify
 CC a patient having or at risk of developing cancer.
 XX
 SQ Sequence 515 AA:
 Query Match 49.8%, Score 100; DB 19; Length 515;
 Best Local Similarity 100.0%; Pred. No. 7.7e-85;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 102 TAVKTIAPPEKKEKNVGLSALICIAEVTSTIEALATPGLPSPMPSTLTMMKNTSELVSSE 164
 Yb 46 TAVKTIAPPEKKEKNVGLSALICIAEVTSTIEALATPGLPSPMPSTLTMMKNTSELVSSE 164
 QY 162 VVLSALAAQKVVDIPREISPYDGLHSQVHLEKITS 201
 Yb 126 VVLSALAAQKVVDIPREISPYDGLHSQVHLEKITS 165
 RESULT 3
 AHC15232
 ID AHC15232 standard; Protein: 1569 AA.
 AC AHC15232;
 DE 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #15232.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX
 XX W0200175067 A2.
 ID 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSRO INC.
 XX
 PI Drmanac RT, Liu C, Tanq JT;

GenCore version 4.5

Copyright. (c) 1993 - 2000 Compuquen Ltd.

OM protein : protein search, using sw model

Run on: July 2, 2002, 16:03.03, Search time: 117.59 Seconds
(without alignments)
189,862 Million cell used

Title: US-09-603-665-5_COPY 1594 1794

perfect score: 201

Sequence: 1 LADPTETFDVIRGLVGNPLP..

Scoring table: OLIQ

Model	Gap	Gapext
Model 1	60.0	60.0
Model 2	60.0	60.0
Model 3	60.0	60.0
Model 4	60.0	60.0
Model 5	60.0	60.0
Model 6	60.0	60.0
Model 7	60.0	60.0
Model 8	60.0	60.0
Model 9	60.0	60.0
Model 10	60.0	60.0
Model 11	60.0	60.0
Model 12	60.0	60.0
Model 13	60.0	60.0
Model 14	60.0	60.0
Model 15	60.0	60.0
Model 16	60.0	60.0
Model 17	60.0	60.0
Model 18	60.0	60.0
Model 19	60.0	60.0
Model 20	60.0	60.0
Model 21	60.0	60.0
Model 22	60.0	60.0
Model 23	60.0	60.0
Model 24	60.0	60.0
Model 25	60.0	60.0
Model 26	60.0	60.0
Model 27	60.0	60.0
Model 28	60.0	60.0
Model 29	60.0	60.0
Model 30	60.0	60.0
Model 31	60.0	60.0
Model 32	60.0	60.0
Model 33	60.0	60.0
Model 34	60.0	60.0
Model 35	60.0	60.0
Model 36	60.0	60.0
Model 37	60.0	60.0
Model 38	60.0	60.0
Model 39	60.0	60.0
Model 40	60.0	60.0
Model 41	60.0	60.0
Model 42	60.0	60.0
Model 43	60.0	60.0
Model 44	60.0	60.0
Model 45	60.0	60.0
Model 46	60.0	60.0
Model 47	60.0	60.0
Model 48	60.0	60.0
Model 49	60.0	60.0
Model 50	60.0	60.0
Model 51	60.0	60.0
Model 52	60.0	60.0
Model 53	60.0	60.0
Model 54	60.0	60.0
Model 55	60.0	60.0
Model 56	60.0	60.0
Model 57	60.0	60.0
Model 58	60.0	60.0
Model 59	60.0	60.0
Model 60	60.0	60.0
Model 61	60.0	60.0
Model 62	60.0	60.0
Model 63	60.0	60.0
Model 64	60.0	60.0
Model 65	60.0	60.0
Model 66	60.0	60.0
Model 67	60.0	60.0
Model 68	60.0	60.0
Model 69	60.0	60.0
Model 70	60.0	60.0
Model 71	60.0	60.0
Model 72	60.0	60.0
Model 73	60.0	60.0
Model 74	60.0	60.0
Model 75	60.0	60.0
Model 76	60.0	60.0
Model 77	60.0	60.0
Model 78	60.0	60.0
Model 79	60.0	60.0
Model 80	60.0	60.0
Model 81	60.0	60.0
Model 82	60.0	60.0
Model 83	60.0	60.0
Model 84	60.0	60.0
Model 85	60.0	60.0
Model 86	60.0	60.0
Model 87	60.0	60.0
Model 88	60.0	60.0
Model 89	60.0	60.0
Model 90	60.0	60.0
Model 91	60.0	60.0
Model 92	60.0	60.0
Model 93	60.0	60.0
Model 94	60.0	60.0
Model 95	60.0	60.0
Model 96	60.0	60.0
Model 97	60.0	60.0
Model 98	60.0	60.0
Model 99	60.0	60.0
Model 100	60.0	60.0

Searched: 747574 seqs, 111073796 residues

Word size : 6

Total number of hits satisfying chosen parameters: 1911

Minimum DB seq length: 0

Maximum PR seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneset 032802.*

1	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1980.DAT
2	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1981.DAT
3	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1982.DAT
4	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1983.DAT
5	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1984.DAT
6	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1985.DAT
7	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1986.DAT
8	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1987.DAT
9	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1988.DAT
10	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1989.DAT
11	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1990.DAT
12	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1991.DAT
13	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1992.DAT
14	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1993.DAT
15	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1994.DAT
16	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1995.DAT
17	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1996.DAT
18	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1997.DAT
19	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1998.DAT
20	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA1999.DAT
21	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA2000.DAT
22	/SIDS1/qcdata/hold_geneseq/geneseqp-emb1/AA2001.DAT

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		% Match					
1	200	99.5	2144	22	AB85029	Protein encoded by	
2	100	49.8	515	19	ABW54099	Homo sapiens BAP28	
3	62	30.8	1569	22	ABG15232	Novel human diapo	
4	8	0	68	22	AWC02460	Novel human diapo	
5	8	4.0	108	22	AAH96435	Putative p. abyss	
6	8	4.0	307	22	AAU14137	Human novel protei	
7	8	4.0	363	21	AB43826	Human cancer assoc	
8	8	4.0	435	22	ABG27718	Novel human diapo	
9	8	4.0	505	22	ARG00754	Novel human diapo	
10	7	3.5	42	22	AAH28751	Peptide #1402 enco	
11	7	3.5	42	22	ABH33938	Peptide #1444 enco	

ALIGNMENT'S

RESULT 1.

RESOLUT
AAB85029

11: AAK85029 standard, Protein; 2144 AA.

AC AAB85029;

06-AUG-2001 (first entry)

XX Protein encoded by BAP28 cDNA consisting of exons 1 to 45
XX

BA128; prostate; cancer, diagnostic; genetic analysis.

XX
US
Homo sapiens.

XX	Key	Location/Qualifiers
FH		

FF	key	local
FF	Misc-difference	1694

	Miscellaneous	Reference	1694		/label - Ser or Asn
Pf					
PtP					

Misc-difference 1854

/label= Ala or Val

FT Misc-difference 1967

Label	Asp	Asn
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1
21	1	1
22	1	1
23	1	1
24	1	1
25	1	1
26	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	1	1
32	1	1
33	1	1
34	1	1
35	1	1
36	1	1
37	1	1
38	1	1
39	1	1
40	1	1
41	1	1
42	1	1
43	1	1
44	1	1
45	1	1
46	1	1
47	1	1
48	1	1
49	1	1
50	1	1
51	1	1
52	1	1
53	1	1
54	1	1
55	1	1
56	1	1
57	1	1
58	1	1
59	1	1
60	1	1
61	1	1
62	1	1
63	1	1
64	1	1
65	1	1
66	1	1
67	1	1
68	1	1
69	1	1
70	1	1
71	1	1
72	1	1
73	1	1
74	1	1
75	1	1
76	1	1
77	1	1
78	1	1
79	1	1
80	1	1
81	1	1
82	1	1
83	1	1
84	1	1
85	1	1
86	1	1
87	1	1
88	1	1
89	1	1
90	1	1
91	1	1
92	1	1
93	1	1
94	1	1
95	1	1
96	1	1
97	1	1
98	1	1
99	1	1
100	1	1

Misc-difference 2017

/label= Gly or Glu

PN
XX
W0200100669-A2.

XX
14-1AN-2111

23-JUN-1983 2000WC-1801183

XX
PF 23-JUN-2000; 2000WC-1801183.

25-1111-1444; 9015-0141223.

18-JAN-2000: 4th

XX.

GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Chen, Xianfeng
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)A
CURRENT APPLICATION NUMBER: US/60/360,039
CURRENT FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 4112
LENGTH: 1133
TYPE: PRT
ORGANISM: Neurospora crassa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(1133)
OTHER INFORMATION: unsure at all xaa locations
US-60-360-039-4112

Query Match: 3.5%; Score 7; DB 7; Length 1133;
Best Local Similarity: 100.0%; Prod No 240002;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SALAAALQ 12
Db 362 SALAAALQ 368

Search completed: July 2, 2002, 16:16:45
Job time: 822 sec

APPLICANT: JONES, Anissa L.
APPLICANT: IRAN, Alanna-Phung B.
APPLICANT: DAHL, Christopher R.
APPLICANT: GIETZEN, Darryl
APPLICANT: CHINN, Joyce
APPLICANT: DUFOUR, Gerard R.
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: YU, Jimmy Y.
APPLICANT: THASON, Olivia
APPLICANT: YAP, Pierre E.
APPLICANT: AMSHEY, Stefan R.
APPLICANT: DAUCHERIVY, Sean C.
APPLICANT: DAM, Tam C.
APPLICANT: LIU, Tommy F.
APPLICANT: NGUYEN, Duy-Viet An
APPLICANT: KLEEFELD, Yoel
APPLICANT: GERSTIN, JR., Edward H.
APPLICANT: PERALLA, Careyna H.
APPLICANT: DAVID, Marie H.
APPLICANT: LEWIS, Samantha A.
APPLICANT: CHEN, Alice J.
APPLICANT: PANZER, Scott R.
APPLICANT: HARRIS, Bernard
APPLICANT: FLORES, Vincent
APPLICANT: MAWARA, Karanish
APPLICANT: LO, Audrey
APPLICANT: LAN, Ruth Y.
APPLICANT: URASHKA, Michael E.
TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
FILE REFERENCE: PI-1231.PCT
CURRENT APPLICATION NUMBER: PCT/US02/09944
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: 60/290,428; 60/291,176
60/291,829; 60/299,428; 60/408,801; 60/299,176
PRIOR FILING DATE: 2001-04-29; 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17.
2001-05-17; 2001-06-19; 2001-06-20; 2001-06-29
NUMBER OF SEQ ID NOS: 792
SOFTWARE: PERL Program
SEQ ID NO 649
LENGTH: 727
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID No: LI:218849.24.or12:2001MAY17
PCT-US02-09944 649

Query Match 3.5% Score 7; DB 1; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TLPHFIS 24
|||||
DB 151 TLPHFIS 157

RESULT 12
US-09-540-2098-7121
; Sequence 7121, Application US/095402098
; GENERAL INFORMATION:
; APPLICANT: Gary L. Herton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540.2098
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 1044
; SEQ ID NO 7121
; LENGTH: 876
; TYPE: PRT
; ORGANISM: B. fragilis
US 09-540-2098-7121

Query Match 3.5% Score 7; DB 5; Length 876;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 NDLEEVG 137
|||||
DB 79 NDLEEVG 85

RESULT 13
US-60-360-039-23646
; Sequence 23646, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Harry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38 10/520521A
; CURRENT APPLICATION NUMBER: US/60/360.039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23646
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-360-039-23646

Query Match 3.5% Score 7; DB 7; Length 883;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ALOKVVE 16
|||||
DB 737 ALOKVVE 743

RESULT 14
US-09-573-655B-1734
; Sequence 1734, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573.655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1734
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1734

Query Match 3.5% Score 7; DB 5; Length 1057;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 TTLAPRV 68
|||||
DB 788 TTLAPRV 794

RESULT 15
US-60-360-039-4112
; Sequence 4112, Application US/60360039

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1001-001

; CURRENT APPLICATION NUMBER: US/09/540,209B

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 10444

; SEQ ID NO 9466

; LENGTH: 417

; TYPE: PRT

; ORGANISM: B.fragilis

US-09-540-209B-9466

Query Match 3.5%, Score 7; DB 5; Length 417;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 NTRLTSL 56

Db 4 NTRLTSL 10

RESULT 7

US-09-540-209B-9330

; Sequence 9330, Application US/09/540,209B

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR

; FILE REFERENCE: 2709.1001-001

; CURRENT APPLICATION NUMBER: US/09/540,209B

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 10444

; SEQ ID NO 9330

; LENGTH: 452

; TYPE: PRT

; ORGANISM: B.fragilis

US-09-540-209B-9330

Query Match 3.5%, Score 7; DB 5; Length 452;

Best Local Similarity 100.0%; Pred. No. 14;02, 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YLLSALA 9

Db 392 YLLSALA 398

RESULT 8

US-60-360-039-5032

; Sequence 5032, Application US/60/360,039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)A

; CURRENT APPLICATION NUMBER: US/60/360,039

; CURRENT FILING DATE: 2002-02-21

; SEQ ID NO 5032

; LENGTH: 488

; TYPE: PRT

; ORGANISM: caenorhabditis elegans

US-60-360-039-5032

Query Match 3.5%, Score 7; DB 7; Length 488,

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 QTEKNWK 85

Db 414 QTEKNWK 420

RESULT 9

US-10-041-018-261

; Sequence 261, Application US/10041018

; GENERAL INFORMATION:

; APPLICANT: Matsuda, Seiichi P.T.

; APPLICANT: Hart, Elizabeth A.

; TITLE OF INVENTION: Diterpene-Producing Unicellular organism

; FILE REFERENCE: P02080051/10025547

; CURRENT APPLICATION NUMBER: US/10/041,018

; CURRENT FILING DATE: 2002-01-07

; PRIOR APPLICATION NUMBER: US 60/259880

; PRIOR FILING DATE: 2001-01-05

; NUMBER OF SEQ ID NOS: 413

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 261

; LENGTH: 579

; TYPE: PRT

; ORGANISM: Zea mays

US-10-041-018-261

Query Match 3.5%, Score 7; DB 6; Length 579;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLSALAA 10

Db 540 LLSALAA 546

RESULT 10

US-60-360-039-6112

; Sequence 6112, Application US/60/360,039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)A

; CURRENT APPLICATION NUMBER: US/60/360,039

; CURRENT FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 6112

; LENGTH: 688

; TYPE: PRT

; ORGANISM: caenorhabditis elegans

US-60-360-039-6112

Query Match 3.5%, Score 7; DB 7; Length 688;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALAALQK 13

Db 500 ALAALQK 506

RESULT 11

PCT-US02-09944-649

; Sequence 649, Application PCT/US0209944

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: DAFFO, Abel

; LOCATION: 1..105
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..105
; OTHER INFORMATION: Ceres Seq. ID 1393053
US-09-620-393B-6607

Query Match 3.5%; Score 7; DB 5; Length 105;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LSALAAL 11
IIIIII
Db 4 LSALAAL 49

RESULT 3
PCT-US02-393B-6606
; Sequence 6606, Application US/09620393B
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1068P
; CURRENT APPLICATION NUMBER: US/09/620,393B
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9948
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 6606
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..121
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..121
; OTHER INFORMATION: Ceres Seq. ID 1393052
US-09-620-393B-6606

Query Match 3.5%; Score 7; DB 5; Length 121;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LSALAAL 11
IIIIII
Db 49 LSALAAL 55

RESULT 4
US-09-620-393B-6605
; Sequence 6605, Application US/09620393B
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1068P
; CURRENT APPLICATION NUMBER: US/09/620,393B
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9948
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 6605
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..158

; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..158
; OTHER INFORMATION: Ceres Seq. ID 1393051
US-09-620-393B-6605

Query Match 3.5%; Score 7; DB 5; Length 158;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LSALAAL 11
IIIIII
Db 86 LSALAAL 92

RESULT 5
PCT-US02-13329-8
; Sequence 8, Application PC/TUS0213329
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: THORNTON, Michael
; APPLICANT: WALLA, Narinder K.
; APPLICANT: GLETZEN, Kimberly J.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: WARREN, Bridget K.
; APPLICANT: RAMKUMAR, Javalazmi
; APPLICANT: YAO, Monique G.
; APPLICANT: JIN, Pei
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: GRAUL, Richard C.
; APPLICANT: YANG, Junming
; APPLICANT: DING, Li
; APPLICANT: FU, Glenn K.
; TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: PI-0415 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/13329
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/287,151; 60/290,514; 60/291,217; 60/314,752;
60/329,217; 60/343,718; 60/343,903
; PRIOR FILING DATE: 2001-04-25; 2001-07-11; 2001-05-15; 2001-08-24;
2001-10-12; 2001-10-19; 2001-11-02
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PRL Program
; SEQ ID NO 8
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7472061C01
PCT-US02-13329-8

Query Match 3.5%; Score 7; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VYLLSAL 8
IIIIII
Db 42 VYLLSAL 48

RESULT 6
US-09-540-209B-9466
; Sequence 9466, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOPHAGES

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 16:16:44, Search time 59.99 seconds
(without alignments)
338 459 Million cell updates/sec

Title: 09-603-665-5_COPY_1754_1954

Perfect score: 201

Sequence: 1 EVVLSALAAALQKVVETLPH

EPYNAIAIAIKKIKETLLE 201

Scoring table:

Gapop 60 0, Gapext 60 0

Searched: 307805 seqs, 10101545 residues

Word size: 6

Total number of hits satisfying chosen parameters: 1089

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Pending Patents AA-New:

- 1: /cgn2_6/ptodata/1/paa/us60_NEW_COMB.pcp.*
- 2: /cgn2_6/ptodata/1/paa/us60_NEW_COMB.pcp.*
- 3: /cgn2_6/ptodata/1/paa/us07_NEW_COMB.pcp.*
- 4: /cgn2_6/ptodata/1/paa/us08_NEW_COMB.pcp.*
- 5: /cgn2_6/ptodata/1/paa/us09_NEW_COMB.pcp.*
- 6: /cgn2_6/ptodata/1/paa/us10_NEW_COMB.pcp.*
- 7: /cgn2_6/ptodata/1/paa/us60_NEW_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8	4.6	543	7	US-09-620-3938-6607	Sequence 20635, A
2	7	3.5	105	5	US-09-620-3938-6607	Sequence 6607, Ap
3	7	3.5	121	5	US-09-620-3938-6606	Sequence 6606, Ap
4	7	3.5	158	5	US-09-620-3938-6605	Sequence 6605, Ap
5	7	3.5	323	1	PCT-US02-13329-8	Sequence 8, Appl1
6	7	3.5	417	5	US-09-540-2098-9466	Sequence 9466, Ap
7	7	3.5	452	5	US-09-540-2098-9330	Sequence 9330, Ap
8	7	3.5	488	7	US-60-360-039-5032	Sequence 5032, Ap
9	7	3.5	579	6	US-10-041-018-261	Sequence 261, App
10	7	3.5	688	7	US-60-360-045-6112	Sequence 6112, Ap
11	7	3.5	727	1	PCT-US02-09044-649	Sequence 649, App
12	7	3.5	876	5	US-09-540-2098-7121	Sequence 7121, Ap
13	7	3.5	883	7	US-60-360-039-23646	Sequence 23646, A
14	7	3.5	1057	5	US-09-573-6558-1734	Sequence 1734, Ap
15	7	3.5	1133	7	US-60-360-039-4112	Sequence 4112, Ap
16	7	3.5	1612	7	US-60-360-045-1628	Sequence 1628, Ap
17	6	3.0	6	7	US-60-380-336-395	Sequence 395, App
18	6	3.0	40	5	US-10-105-290-4818	Sequence 4818, Ap
19	6	3.0	41	5	US-09-620-3938-1761	Sequence 1761, Ap
20	6	3.0	55	5	US-10-155-881-9573	Sequence 9573, Ap
21	6	3.0	57	5	US-09-620-3938-550	Sequence 550, App
22	6	3.0	62	6	US-10-155-881-21249	Sequence 21239, A
23	6	3.0	64	6	US-10-155-881-9610	Sequence 9610, Ap
24	6	3.0	73	5	US-10-002-344A-246	Sequence 240, App
25	6	3.0	73	5	US-10-002-344A-241	Sequence 241, App
26	6	3.0	85	6	US-10-137-337-730	Sequence 730, App

ALIGNMENTS

RESULT 1
US-60-360-039-20635
; Sequence 20635, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PREVENTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)A

; CURRENT APPLICATION NUMBER: US/60/360,039

; CURRENT FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 20635

; LENGTH: 543

; TYPE: PRT

; ORGANISM: Rhodospseudomonas palustris

US-60-360-039-20635

Query Match 4.0%; Score 8; DB 7; Length 543;
Best Local Similarity 100.0%; Pctd. No. 12;
Matches 4; Conservative 0; Mismatches 0; Gaps 0;

QY 4 LLSALAAAL 11

DB 177 LLSALAAAL 184

RESULT 2

US-09-620-3938-6607

; Sequence 6607, Application US/09620393B

; GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID

; TITLE OF INVENTION: THERREY

; FILE REFERENCE: 2750-1068P

; CURRENT APPLICATION NUMBER: US/09/620,393B

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 9948

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 6607

; LENGTH: 105

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: misc_feature

US-09-724-059-1235861

Query Match
Best Local Similarity 100.0%; Score 8; DB 21; Length 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 KEELTSHQ 111
|||||||
DB 13 KEELTSHQ 20

RESULT 14
US-09-724-059-1235861
; Sequence 1235861, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark
; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500DSM
; CURRENT APPLICATION NUMBER: US/09/724-059
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: MacAllister
; SEQ ID NO 1235861
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide sequence
US-09-724-059-1235861

Query Match
Best Local Similarity 100.0%; Score 8; DB 21; Length 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 KEELTSHQ 111
|||||||
DB 13 KEELTSHQ 20

RESULT 14
US-09-724-059-1235861
; Sequence 1235861, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark
; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500DSM
; CURRENT APPLICATION NUMBER: US/09/724-059
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: MacAllister
; SEQ ID NO 1235861
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide sequence
US-09-724-059-1235861

Query Match
Best Local Similarity 100.0%; Score 8; DB 21; Length 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 KEELTSHQ 111
|||||||
DB 13 KEELTSHQ 20

RESULT 15
US-09-724-059-1236203
; Sequence 1236203, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark
; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500DSM
; CURRENT APPLICATION NUMBER: US/09/724-059
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: MacAllister
; SEQ ID NO 1236203
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide sequence
US-09-724-059-1236203

Query Match
Best Local Similarity 100.0%; Score 8; DB 21; Length 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 KEELTSHQ 111
|||||||
DB 13 KEELTSHQ 20

Search completed: July 2, 2002 16:14:35
Job time: 692 sec

```
Query Match 35.8%; Score 72; DB 21; Length 135;
Best Local Similarity 100.0%; Pred. No. 1, 50, 62;
Matches 72, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 13 FVETLPHFISPYLEGIISQVHLEKITSEMGSASQANIRLTSIKKTLATTLAPRVLLPA 72
      |||
DB 44 KVETLPHFISPYLEGIISQVHLEKITSEMGSASQANIRLTSIKKTLATTLAPRVLLPA 103

QY 73 IKTKYQIEKNW 84
      |||
DB 104 IKTKYQIEKNW 115

RESULT 9
US-09-757-028-2541
: Sequence 2541, Application US/09757028
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PM001
: CURRENT APPLICATION NUMBER: US/09/757,028
: PRIOR FILING DATE: 2001-01-09
: PRIOR APPLICATION NUMBER: 60/179,065
: PRIOR FILING DATE: 2000-01-31
: PRIOR APPLICATION NUMBER: 60/180,628
: PRIOR FILING DATE: 2000-02-04
: NUMBER OF SEQ ID NOS: 2650
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2541
: LENGTH: 105
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (18)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (78)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (89)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (100)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-757-028-2541

Query Match 29.4%; Score 59; DB 21; Length 105;
Best Local Similarity 100.0%; Pred. No. 9, 20, 50;
Matches 59, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 15 VETLPHFISPYLEGIISQVHLEKITSEMGSASQANIRLTSIKKTLATTLAPRVLLPAI 73
      |||
DB 19 VETLPHFISPYLEGIISQVHLEKITSEMGSASQANIRLTSIKKTLATTLAPRVLLPAI 77

RESULT 10
US-09-724-059-1234367
: Sequence 1234367, Application US/09724059
: GENERAL INFORMATION:
: APPLICANT: Choo, Yen
: APPLICANT: Klug, Aaron
: APPLICANT: Isalan, Mark
: TITLE OF INVENTION: Nucleic Acid Binding Proteins
: FILE REFERENCE: P25000SM
: CURRENT APPLICATION NUMBER: US/09/724,059
: PRIOR FILING DATE: 2000-11-28
: CURRENT APPLICATION NUMBER: US/09/724,059
: PRIOR FILING DATE: 1997-05-23
: NUMBER OF SEQ ID NOS: 1407122
: OTHER INFORMATION: Polypeptide sequence
```

```
: SOFTWARE: MacAllister
: SEQ ID NO 1234367
: LENGTH: 23
: TYPE: PRT
: ORGANISM: Artificial sequence
: FEATURE:
: OTHER INFORMATION: Polypeptide sequence
US-09-724-059-1234367

Query Match 4.0%; Score 8; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 7, 3;
Matches 8, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 104 KEELTSHQ 111
      |||
DB 13 KEELTSHQ 20

RESULT 11
US-09-724-059-1234745
: Sequence 1234745, Application US/09724059
: GENERAL INFORMATION:
: APPLICANT: Choo, Yen
: APPLICANT: Klug, Aaron
: APPLICANT: Isalan, Mark
: TITLE OF INVENTION: Nucleic Acid Binding Proteins
: FILE REFERENCE: P25000SM
: CURRENT APPLICATION NUMBER: US/09/724,059
: PRIOR FILING DATE: 2000-11-28
: PRIOR APPLICATION NUMBER: GB9710809 6
: PRIOR FILING DATE: 1997-05-23
: NUMBER OF SEQ ID NOS: 1407122
: SOFTWARE: MacAllister
: SEQ ID NO 1234745
: LENGTH: 23
: TYPE: PRT
: ORGANISM: Artificial sequence
: FEATURE:
: OTHER INFORMATION: Polypeptide sequence
US-09-724-059-1234745

Query Match 4.0%; Score 8; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 7, 3;
Matches 8, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 104 KEELTSHQ 111
      |||
DB 13 KEELTSHQ 20

RESULT 12
US-09-724-059-1235105
: Sequence 1235105, Application US/09724059
: GENERAL INFORMATION:
: APPLICANT: Choo, Yen
: APPLICANT: Klug, Aaron
: APPLICANT: Isalan, Mark
: TITLE OF INVENTION: Nucleic Acid Binding Proteins
: FILE REFERENCE: P25000SM
: CURRENT APPLICATION NUMBER: US/09/724,059
: PRIOR FILING DATE: 2000-11-28
: PRIOR APPLICATION NUMBER: GB9710809 6
: PRIOR FILING DATE: 1997-05-23
: NUMBER OF SEQ ID NOS: 1407122
: SOFTWARE: MacAllister
: SEQ ID NO 1235105
: LENGTH: 23
: TYPE: PRT
: ORGANISM: Artificial sequence
: FEATURE:
: OTHER INFORMATION: Polypeptide sequence
```


LOCATION: 1967
 OTHER INFORMATION: Xaa-Asp or Asn
 NAME/KEY: VARIANT
 LOCATION: 2017
 OTHER INFORMATION: Xaa-Gly or Glu
 US-09-603-665-5

Query Match 99.5%, Score 200; DB 26; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 3.3e-188;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVLLSALAAQKQVVEIHPHISPYLEGILSOVHLEKIIISMGSSASQANILSLKKTL 1813
 DB 1754 EYVLLSALAAQKQVVEIHPHISPYLEGILSOVHLEKIIISMGSSASQANILSLKKTL 1813
 QY 61 ATTAPRVLLPAIKTYKQIEYNNWNNHMGFPMSILGEHIGXMKPEELISHSQLTAFLE 120
 DB 1814 ATTAPRVLLPAIKTYKQIEYNNWNNHMGFPMSILGEHIGXMKPEELISHSQLTAFLE 1873
 QY 121 ALDFRAQSENDELEEVGKTFENIICVLAMVVKLSVTERPFLFFKLFQWAKTEDAPKDEL 180
 DB 1874 ALDFRAQSENDELEEVGKTFENIICVLAMVVKLSVTERPFLFFKLFQWAKTEDAPKDEL 1933
 QY 181 LTFYNLADCIAEKLGLETLF 201
 DB 1934 LTFYNLADCIAEKLGLETLF 1954

RESULT 2
 US-60-176-880-4
 : Sequence 4; Application US/60174880
 : GENERAL INFORMATION:
 : APPLICANT: Barry, Caroline
 : APPLICANT: Bouquelier, Lydie
 : APPLICANT: Choumakov, Ilya
 : TITLE OF INVENTION: A NOVEL BAP28 GENE AND PROTEIN.
 : FILE REFERENCE: 67 US2, PRO
 : CURRENT APPLICATION NUMBER: US760/176,880
 : CURRENT FILING DATE: 2000-01-18
 : EARLIER APPLICATION NUMBER: US 60/141,324
 : EARLIER FILING DATE: 1999-06-25
 : NUMBER OF SEQ ID NOS: 12
 : SOFTWARE: Patent.pro
 : SEQ ID NO 4
 : LENGTH: 2144
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: VARIANT
 : LOCATION: 1694
 : OTHER INFORMATION: Xaa-Ser or Asn
 : FEATURE:
 : NAME/KEY: VARIANT
 : LOCATION: 1854
 : OTHER INFORMATION: Xaa-Ala or Val
 : FEATURE:
 : NAME/KEY: VARIANT
 : LOCATION: 1967
 : OTHER INFORMATION: Xaa-Asp or Asn
 : FEATURE:
 : NAME/KEY: VARIANT
 : LOCATION: 2017
 : OTHER INFORMATION: Xaa-Gly or Glu
 US-60-176-880-4

Query Match 99.5%, Score 200; DB 26; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 3.3e-188;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVLLSALAAQKQVVEIHPHISPYLEGILSOVHLEKIIISMGSSASQANILSLKKTL 60
 DB 1754 EYVLLSALAAQKQVVEIHPHISPYLEGILSOVHLEKIIISMGSSASQANILSLKKTL 1813

DB 1754 EYVLLSALAAQKQVVEIHPHISPYLEGILSOVHLEKIIISMGSSASQANILSLKKTL 1813
 QY 61 ATTAPRVLLPAIKTYKQIEYNNWNNHMGFPMSILGEHIGXMKPEELISHSQLTAFLE 120
 DB 1814 ATTAPRVLLPAIKTYKQIEYNNWNNHMGFPMSILGEHIGXMKPEELISHSQLTAFLE 1873
 QY 121 ALDFRAQSENDELEEVGKTFENIICVLAMVVKLSVTERPFLFFKLFQWAKTEDAPKDEL 180
 DB 1874 ALDFRAQSENDELEEVGKTFENIICVLAMVVKLSVTERPFLFFKLFQWAKTEDAPKDEL 1933
 QY 181 LTFYNLADCIAEKLGLETLF 201
 DB 1934 LTFYNLADCIAEKLGLETLF 1954

RESULT 3
 US-08-936-487-52
 : Sequence 52; Application US/08936487
 : GENERAL INFORMATION:
 : APPLICANT: Howcock, Anne M.
 : APPLICANT: Baer, Richard
 : TITLE OF INVENTION: Compositions and Methods Comprising
 : TITLE OF INVENTION: BARD1 and other BRCA1 Binding Proteins
 : NUMBER OF SEQUENCES: 130
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Arnold, White & Durkee
 : STREET: P.O. Box 4433
 : CITY: Houston
 : STATE: Texas
 : COUNTRY: USA
 : ZIP: 77210
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.10
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/936,487
 : FILING DATE: Concurrently Herewith
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 00/025,296
 : FILING DATE: 20-SEP-1996
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 00/042,611
 : FILING DATE: 03-APR-1997
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 00/042,985
 : FILING DATE: 04-APR-1997
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Hibler, David W.
 : REGISTRATION NUMBER: P-41,071
 : REFERENCE/POCKET NUMBER: UTSD:499
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 512/418-3000
 : TELEFAX: 512/474-7577
 : INFORMATION FOR SEQ ID NO: 52:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 515 amino acids
 : TYPE: amino acid
 : STRANDEDNESS:
 : TOPOLOGY: linear
 US-08-936-487-52

Query Match 49.8%, Score 100; DB 13; Length 515;
 Best Local Similarity 100.0%; Pred. No. 1e-89;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVLLSALAAQKQVVEIHPHISPYLEGILSOVHLEKIIISMGSSASQANILSLKKTL 60
 DB 1754 EYVLLSALAAQKQVVEIHPHISPYLEGILSOVHLEKIIISMGSSASQANILSLKKTL 1813

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

GM protein - protein search, using sw model

Run on: July 2, 2002, 16:34:34 : Search time 59.3 seconds
(without alignments)
140,012 Million cell updates/sec

Title: US-09-603-665-5_COPY_1754_1954

Perfect score: 201

Sequence: 1 KWLISALAAAKQVETLPH.....TFYNLADCIARKKLEPTLF 201

Scoring table: oligo

Gapop 60.0 , Gapext 60.0

Searched: 3502263 seqs, 35190561 residues

Word size: 6

Total number of hits satisfying chosen parameters: 4229

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Pending_Patents_AA_Main:*

- 1: /cqn2_6/prodata/2/paa/us086_COMB.pep.*
- 2: /cqn2_6/prodata/2/paa/us086_COMB.pep.*
- 3: /cqn2_6/prodata/2/paa/us087_COMB.pep.*
- 4: /cqn2_6/prodata/2/paa/us087_COMB.pep.*
- 5: /cqn2_6/prodata/2/paa/us081_COMB.pep.*
- 6: /cqn2_6/prodata/2/paa/us082_COMB.pep.*
- 7: /cqn2_6/prodata/2/paa/us083_COMB.pep.*
- 8: /cqn2_6/prodata/2/paa/us084_COMB.pep.*
- 9: /cqn2_6/prodata/2/paa/us085_COMB.pep.*
- 10: /cqn2_6/prodata/2/paa/us086_COMB.pep.*
- 11: /cqn2_6/prodata/2/paa/us087_COMB.pep.*
- 12: /cqn2_6/prodata/2/paa/us088_COMB.pep.*
- 13: /cqn2_6/prodata/2/paa/us089_COMB.pep.*
- 14: /cqn2_6/prodata/2/paa/us090_COMB.pep.*
- 15: /cqn2_6/prodata/2/paa/us091_COMB.pep.*
- 16: /cqn2_6/prodata/2/paa/us092_COMB.pep.*
- 17: /cqn2_6/prodata/2/paa/us093_COMB.pep.*
- 18: /cqn2_6/prodata/2/paa/us094_COMB.pep.*
- 19: /cqn2_6/prodata/2/paa/us095_COMB.pep.*
- 20: /cqn2_6/prodata/2/paa/us096_COMB.pep.*
- 21: /cqn2_6/prodata/2/paa/us097_COMB.pep.*
- 22: /cqn2_6/prodata/2/paa/us098_COMB.pep.*
- 23: /cqn2_6/prodata/2/paa/us099_COMB.pep.*
- 24: /cqn2_6/prodata/2/paa/us100_COMB.pep.*
- 25: /cqn2_6/prodata/2/paa/us101_COMB.pep.*
- 26: /cqn2_6/prodata/2/paa/us086_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	200	99.5	2144	20	US-09-603-665-5 Sequence 5, Appli
2	200	99.5	2144	26	US-60-176-880-4 Sequence 4, Appli
3	100	49.8	515	13	US-08-646-487-52 Sequence 52, Appl
4	100	49.8	515	26	US-60-043-611-52 Sequence 52, Appl
5	100	49.8	515	26	US-60-643-685-52 Sequence 52, Appl
6	100	49.8	1149	26	US-60-339-453-167 Sequence 167, App
7	100	49.8	2144	26	US-60-141-323-3 Sequence 3, Appli

8	72	35.8	135	21	US-09-757-028-2590	Sequence 2590, Ap
9	59	29.4	105	21	US-09-757-028-2541	Sequence 2541, Ap
10	8	4.0	23	21	US-09-724-059-124347	Sequence 124347,
11	8	4.0	23	21	US-09-724-059-124745	Sequence 124745,
12	8	4.0	23	21	US-09-724-059-1235185	Sequence 1235185,
13	8	4.0	23	21	US-09-724-059-1235483	Sequence 1235483,
14	8	4.0	23	21	US-09-724-059-1235861	Sequence 1235861,
15	8	4.0	23	21	US-09-724-059-1236203	Sequence 1236203,
16	8	4.0	23	21	US-09-724-059-1236617	Sequence 1236617,
17	8	4.0	23	21	US-09-724-059-1236959	Sequence 1236959,
18	8	4.0	23	21	US-09-724-059-1237319	Sequence 1237319,
19	8	4.0	23	21	US-09-724-059-1237697	Sequence 1237697,
20	8	4.0	23	21	US-09-724-059-1242341	Sequence 1242341,
21	8	4.0	23	21	US-09-724-059-1242593	Sequence 1242593,
22	8	4.0	23	21	US-09-724-059-1242833	Sequence 1242833,
23	8	4.0	23	21	US-09-724-059-1243085	Sequence 1243085,
24	8	4.0	23	21	US-09-724-059-1243337	Sequence 1243337,
25	8	4.0	23	21	US-09-724-059-1243555	Sequence 1243555,
26	8	4.0	23	21	US-09-724-059-1243841	Sequence 1243841,
27	8	4.0	23	21	US-09-724-059-1244069	Sequence 1244069,
28	8	4.0	23	21	US-09-724-059-1244369	Sequence 1244369,
29	8	4.0	23	21	US-09-724-059-1244561	Sequence 1244561,
30	8	4.0	23	21	US-09-724-059-1245676	Sequence 1245676,
31	8	4.0	23	21	US-09-724-059-1248196	Sequence 1248196,
32	8	4.0	23	21	US-09-724-059-1270596	Sequence 1270596,
33	8	4.0	23	21	US-09-724-059-1273116	Sequence 1273116,
34	8	4.0	23	21	US-09-724-059-1275636	Sequence 1275636,
35	8	4.0	23	21	US-09-724-059-1277916	Sequence 1277916,
36	8	4.0	23	21	US-09-724-059-1280676	Sequence 1280676,
37	8	4.0	23	21	US-09-724-059-1282956	Sequence 1282956,
38	8	4.0	23	21	US-09-724-059-1283356	Sequence 1283356,
39	8	4.0	23	21	US-09-724-059-1283796	Sequence 1283796,
40	8	4.0	23	21	US-09-724-059-128409	Sequence 128409,
41	8	4.0	23	21	US-09-724-059-128445	Sequence 128445,
42	8	4.0	23	21	US-09-724-059-128894	Sequence 128894,
43	8	4.0	23	21	US-09-724-059-129213	Sequence 129213,
44	8	4.0	23	21	US-09-724-059-129369	Sequence 129369,
45	8	4.0	23	21	US-09-724-059-129393	Sequence 129393,

ALIGNMENTS

RESULT 1
US-09-603-665-5
Sequence 5, Application US/09603665
GENERAL INFORMATION:
APPLICANT: Barry, Caroline
APPLICANT: Hougholter, Lydie
APPLICANT: Chumakov, Iliya
TITLE OF INVENTION: A N-VEL RAF28 GENE AND PROTEIN
FILE REFERENCE: GENSET-063A05
CURRENT APPLICATION NUMBER: US/09/603,665
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/141,323
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 60/176,880
PRIOR FILING DATE: 2000-01-18
NUMBER OF SEQ. ID NOS: 63
SOFTWARE: Patent.pm
SEQ. ID NO 5
LENGTH: 2144
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: 1694
OTHER INFORMATION: Xaa-Ser or Asn
NAME/KEY: VARIANT
LOCATION: 1854
OTHER INFORMATION: Xaa-Ala or Val
NAME/KEY: VARIANT